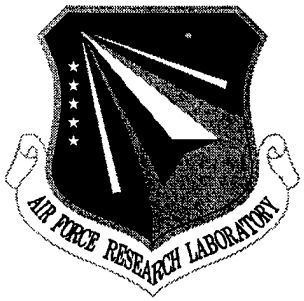
AFRL-IF-RS-TR-1998-187
Final Technical Report
November 1998



USE OF A'SCAPE FOR THE DETECTION OF LUNG TUMORS

Research Associates for Defense Conversion, Inc.

Mohamed A. Slamani

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-1998-187 has been reviewed and is approved for publication.

APPROVED:

DAVID D. FERRIS, JR.
Project Engineer

FOR THE DIRECTOR:

JOSEPH CAMERA, Deputy Chief
Information & Intelligence Exploitation Division
Information Directorate

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/IFEA, 26 Electronic Pky, Rome, NY 13441-4514. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED	
	November 1998	Final	May 97 - Apr 98

4. TITLE AND SUBTITLE

USE OF A'SCAPE FOR DETECTION OF LUNG TUMORS

5. FUNDING NUMBERS

C - F30602-95-C-0121
PE - 61102F
PR - 2304
TA - E8
WU - PE

6. AUTHOR(S)

Mohamed A. Slamani

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Research Associates for Defense Conversion, Inc.
10002 Hillside Terrace
Marcy NY 13403

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFEA
26 Electronic Pky
Rome NY 13441-4514

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-1998-187

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: David D. Ferris/IFEA/(315) 330-4408

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The Automatic Statistical Characterization and Partitioning of Environment (A'SCAPE) has previously been successfully used to characterize infrared (IR) and radar land scenes as a target predetection state (DoD application) and, in another application (law enforcement), to detect weapons concealed underneath clothing in scenes collected by different types of sensors including IR and Millimeter-wave sensors.

In this dual-use technology project, the prime emphasis was placed on structuring A'SCAPE to (1) detect tumors in lung tissue, and (2) classify a particular tumor as being either benign or malignant using computer tomography (CT) data. It is shown that A'SCAPE (1) can successfully highlight suspected tumor tissue within the (CT) lung image, and (2) has the potential to detect and classify tumors. The latter point can be proven however only if more data of both benign and malignant tumors is made available. During this effort, a library of only 5 usable data sets were available. Two of these sets are used to develop rules and the remaining three are used to test the rules. However, the results demonstrate promise for future application of this approach.

14. SUBJECT TERMS

Target Predetection stage, tumor detection, concealed weapons detection

15. NUMBER OF PAGES

90

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Table of Content

Acknowledgements

This research was supported by AFRL Rome Research Site in Rome, NY under contract number F30602-95-C-0121.

I would like to thank Dr. Claudia Henschke from the New-York Hospital-Cornell Medical Center, New-York, NY, for her continuous help and support in sending the medical data processed in this report and in evaluating the processing.

Chapter I

Introduction

One of the most challenging problems in the Medical field today is the early detection of tumors. Whenever a cancerous tumor is not detected in time, its treatment becomes difficult if not impossible. In an attempt to assist in the solution of the problem, a signal processing procedure referred to as the Automatic Statistical Characterization And Partitioning of Environment (A'SCAPE) [1] was utilized. A'SCAPE was successfully used to characterize infrared (IR) and radar land scenes as a target pre-detection stage (DoD application) [1], and, in another application (law enforcement), to detect weapons concealed underneath clothing in scenes collected by different types of sensors including IR and Millimeter-wave sensors [1].

In this project, the prime emphasis was placed on structuring A'SCAPE to (1) detect tumors in lung tissue, and (2) classify a particular tumor as being either benign or malignant using computed tomography (CT) data. It is shown that A'SCAPE (1) can successfully highlight suspected tumor tissue within the (CT) lung image, and (2) has the potential to detect and classify tumors. The latter point can be proven however only if more data of both benign and malignant tumors is made available. During this effort, a library of only 5 usable data sets were available. Two of these sets are used to develop rules and the remaining three are used to test the rules. However, the results demonstrate promise for future application of this approach.

The A'SCAPE procedure is presented in Chapter II. In Chapter III, A'SCAPE is applied to 5 different cases of cancerous lungs. Processing and results are provided in the same chapter. A conclusion and a set of recommendations for future work are given in chapter IV.

Chapter II

A'SCAPE Procedure

2.1 - Introduction

The goal of this work is to enhance and detect tumors in the data collected by CT scans. Enhancement of the tumor area constitutes a de-cluttering problem where the objects other than the tumors are considered as clutter. In this work, two methods are used for this purpose. First, based on the difference between the average power levels of the different regions in a scene, a mapping procedure [2] is used to isolate the homogeneous regions. This stage is needed to clean the image from unneeded isolated cells and to determine the regions that can be suspected of containing tumors. The procedure has the potential to separate between regions of different average power levels even when the power levels are very close. On the other hand, for those regions that are contiguous and non-homogeneous but with similar average power levels, a statistical procedure [3] is used to separate them. The procedure is a state of the art method that groups data with similar statistical distributions into subsets called regions. By doing so, different regions are obtained; each with a corresponding probability density function (PDF) that would best approximate the statistical distribution of the data in the corresponding region. Each group of data defines a homogeneous region. Both, the Mapping and Statistical procedures are part of the Automated Statistical Characterization and Partitioning of the Environment (A'SCAPE) [1] shown in Figure 2.1.

Both, the Mapping and Statistical procedures are presented next.

2.2 - Mapping Procedure

Assuming that a scene consists of a set of patches, let LP refer to the patch with the lowest average magnitude and let also RPs refer to the remaining patches. Using the fact that RPs patches, on average, have stronger magnitudes, the mapping procedure

begins by setting a threshold that results in a specified fraction of LP pixels. Image processing is then used to establish the LP and RPs patches. If the final image contains a significantly different fraction of LP than originally established by the initial threshold, the process is repeated with a new threshold. The mapping procedure iterates until it is satisfied that the final scene is consistent with the previous specified threshold. Finally, edges of all patches are detected using an image processing technique referred to as the "unsharp masking".

As a result and as shown in Figure 2.2, the mapping procedure consists of two stages. In the first stage, the patch with the lowest average power, LP, among all remaining patches, RPs is identified. In the second stage, edges of the LP are enhanced and detected.

These two stages are repeated to identify the next LP and so on. The mapping procedure is repeated continuously until it is not possible to further separate between patches, and all patches are declared to be homogeneous. Once all patches have been found, every patch is processed by the mapping procedure, as discussed above, for detection of subpatches.

The Mapping procedure as described in [2] is modified to simultaneously compute all thresholds that would separate between the identified regions in a scene. As shown in Figure 2.3, this step is performed by the block referred to as the Automatic Thresholds Computation (ATC). Then, for each threshold the Mapping procedure: (1) quantizes the scene at that threshold, (2) extracts the different regions through a low pass filter, and (3) extracts the edges corresponding to the different regions using a high pass filter. The resulting scene is referred to as the component image. These steps are repeated for each threshold. At the end, the component images are combined to form a composite image which will be equivalent to the original image but would have the different regions enhanced and represented visually by different colors. One or more of these regions would represent tumor(s). Note that if the data in the original scene consists of 8 bits, it then contains 256 color levels. On the other hand, the composite image consists of only N color levels where N is the number of thresholds found by the ATC stage.

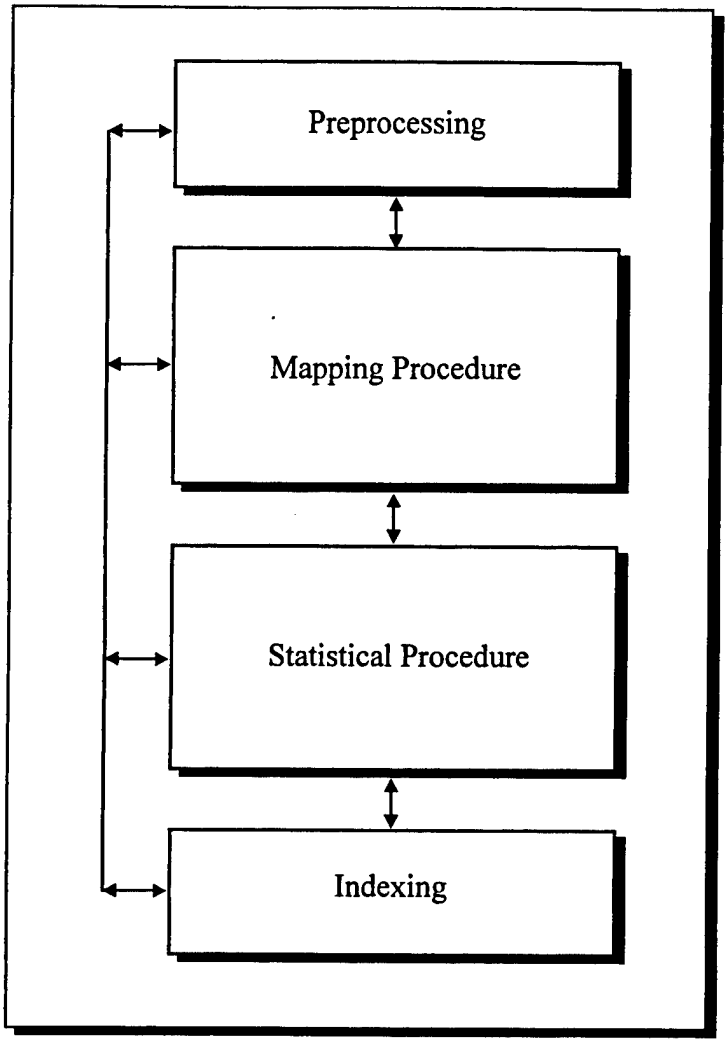


Figure 2.1 – Block Diagram of A'SCAPE

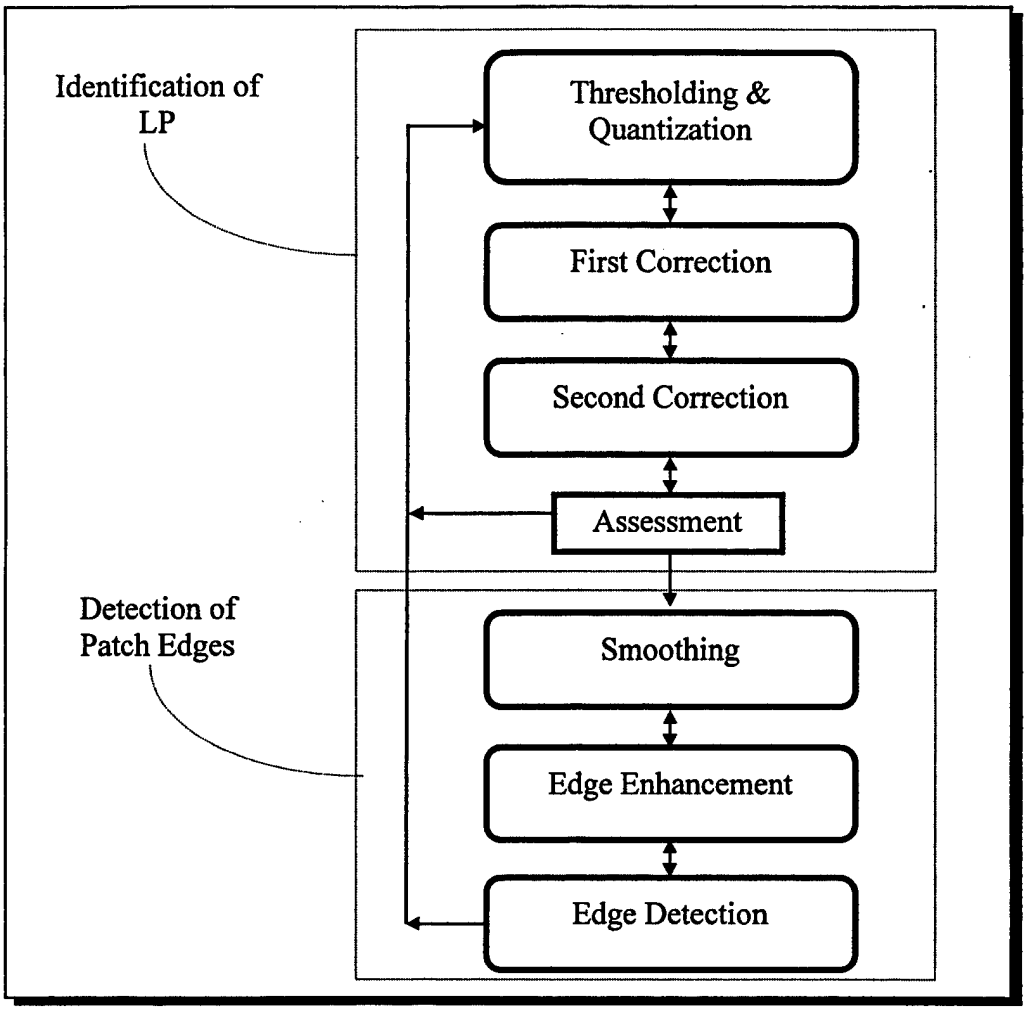


Figure 2.2 – Block Diagram of the Mapping Procedure

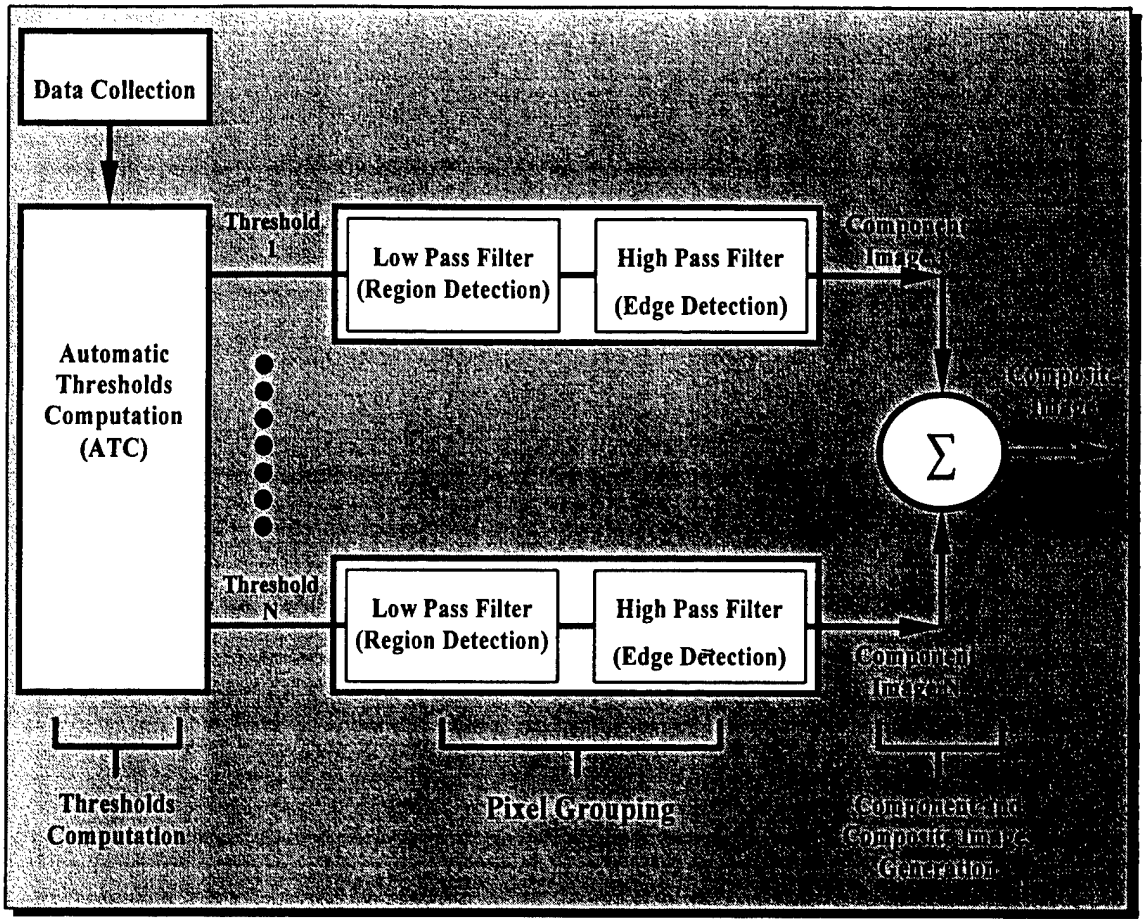


Figure 2.3 - Block Diagram of the Modified Mapping Procedure

2.3 - Statistical Procedure

2.3.1 - Introduction

When it is not possible to separate between contiguous non-homogeneous regions based on power levels, separation between the regions can be obtained using the data distribution in each region. Specifically, using the Statistical procedure of A'SCAPE [3], one can investigate the PDF(s) that can approximate the data in the scene and based on the outcome decide whether the scene is homogeneous. And, if it is not, determine the different homogeneous regions and their boundaries. The Statistical procedure of A'SCAPE is based on the Ozturk Algorithm which is presented next followed by the details of the Statistical Procedure.

2.3.2 - Ozturk Algorithm [4]

In signal processing applications it is common to assume a Gaussian problem in the design of optimal signal processors. However, non-Gaussian processes do arise in many situations. When the possibility of a non-Gaussian problem is encountered, the question as to which probability distributions should be utilized in a specific situation for modeling the data needs to be answered. In practice, the underlying probability distributions are not known a priori and, therefore, have to be determined experimentally by analyzing the random data.

Classical techniques fail to approximate the underlying probability distributions for a given set of data. They only perform goodness-of-fit tests by providing an answer to the question "Is the set of random data statistically consistent with a specified distribution to within a desired confidence level ?". Furthermore, these techniques require a large number of samples (typically several thousands) to perform the test.

On the other hand, the Ozturk algorithm is a new statistical algorithm capable of approximating the probability distribution function (PDF) of a set of random data using only 100 statistically independent sample points. It consists of two modes: the goodness-of-fit test mode and the PDF approximation mode. The first mode determines whether a sample data set is statistically consistent with a pre-specified PDF. The second mode

selects the 'best' approximate PDF from a variety of PDFs and is simply an extension of the goodness-of-fit test.

The two modes are described in detail next.

2.3.2.a - Goodness-of-fit Test Mode

The goodness-of-fit test is an empirical algorithm which determines if the sample data is statistically consistent with a given distribution, called the null hypothesis, to within a desired confidence level. In the Ozturk algorithm, linked vectors are constructed for both the null hypothesis and the sample data set. The confidence contours are constructed around the terminal point of the null hypothesis linked vector.

2.3.2.b - The Linked Vector

To obtain the linked vectors, consider:

1 - the sample data set: $x_1, x_2, x_3, \ldots, x_N$ with sample mean μ_x, sample standard deviation σ_x, and length N,

2 - a null hypothesis data set generated from any available distribution against which the sample set will be tested: $z_1, z_2, z_3, \ldots, z_N$ with zero mean, unit variance and length N, and

3 - an auxiliary data set generated from the standard Gaussian: $w_1, w_2, w_3, \ldots, w_N$ with zero mean, unit variance and length N.

Next, reorder all data sets (ordered statistics) with the smallest value first:

$$x_{1:N}, x_{2:N}, x_{3:N}, \ldots, x_{N:N}$$

$$z_{1:N}, z_{2:N}, z_{3:N}, \ldots, z_{N:N}$$

$$w_{1:N}, w_{2:N}, w_{3:N}, \ldots, w_{N:N}$$

Let, $y_{i:N}$, for the sample linked vector be defined as:

$$y_{i:N} = \frac{x_{i:N} - \mu_x}{\sigma_x} \tag{1}$$

The magnitude of the sample linked vector is the absolute value of $y_{i:N}$. Also, let $t_{i:N}$, for the null hypothesis be defined as the expected value of the i^{th} ordered statistic of the null hypothesis distribution:

$$t_{i:N} = E[z_{i:N}]. \quad (2)$$

The magnitude of the null hypothesis linked vector is the absolute value of $t_{i:N}$. Finally, let $m_{i:N}$, be the expected value of the i^{th} ordered statistic of the auxiliary distribution (the Gaussian distribution in this case):

$$m_{i:N} = E[w_{i:N}]. \quad (3)$$

The expected values are obtained through a Monte-Carlo simulation consisting of 2000 generated data sets for both the null hypothesis and auxiliary data sets.

The set of angles associated with each linked vector is defined as:

$$\theta_i = \pi \phi(m_{i:N}) \quad (4)$$

where:

$$\phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp(-\frac{t^2}{2}) dt. \quad (5)$$

Thus, the angle depends on the value of the Gaussian distribution function, evaluated at each expected value of the reference distribution ordered statistic.

Next, set up the coordinate system $Q_k=[u_k,v_k]$, where:

$$u_k = \frac{1}{k} \sum_{i=1}^{k} |y_{i:N}| \cos\theta_i; \quad k = 1,2,3,\ldots,N \quad (6)$$

$$v_k = \frac{1}{k} \sum_{i=1}^{k} |y_{i:N}| \sin\theta_i; \quad k = 1,2,3,\ldots,N \quad (7)$$

and

$$Q_o=[u_o,v_o]=(0,0), \quad (8)$$

for the sample linked vector. The null hypothesis linked vector is obtained by replacing $y_{i:N}$ with $t_{i:N}$ in the expressions for u_k and v_k above.

Note that the angle θ_i is solely dependent on the auxiliary distribution for all linked vectors, while the magnitude is solely dependent on the data chosen for the linked vector for the sample data and null hypothesis.

Furthermore note that $y_{i:N}$ and $t_{i:N}$ are ordered statistics from smallest to largest, while the magnitudes of $y_{i:N}$ and $t_{i:N}$ are no longer true ordered statistics, due to standardization. Since $y_{i:N}$ and $t_{i:N}$ contain negative values due to standardization, then their magnitudes would begin large, decrease to approximately zero and then increase again.

Also, when the length N of the sample data set is large (on the order of 50 points), then the linked vector is a smooth arc.

2.3.2.c - The Confidence Contours

The linked vector for the null hypothesis is based on the expected values of the ordered statistic $z_{i:N}$ for the 2000 Monte-Carlo simulations. Thus if one considers just one point along the linked vector, in particular the end point, the Monte-Carlo simulation provides 2000 points of which only the expected value is plotted,. However, these 2000 points can also be analyzed for their distribution.

The confidence contours associated with the end point of the null hypothesis are determined by fitting a three dimensional bell shaped (bivariate Gaussian) curve to the 2000 points arising from the distribution of the Monte-Carlo end points for the null hypothesis linked vector. The contours of constant density of this distribution are then plotted for various values of the parameter alpha, (e.g., 0.01, 0.05, and 0.10), where alpha is the probability that the end point falls outside the specified contour given that the data is from the null hypothesis distribution. Then unity minus alpha is known as the confidence contour. Alpha is known as the significance level of the test. Also, 1-Alpha, is referred to as the confidence level of the test.

This may be repeated for any N points of the ordered statistic, $z_{i:N}$, along the null hypothesis linked vector. If the sample data is truly consistent with the null hypothesis, then the sample data's linked vector trajectory will pass through a series of hoops defined by the confidence contours. Because the human eye can readily detect whether or not the linked vectors are closely following the same trajectory, only the last set of confidence contours are typically used.

As the significance level of the test increases, the corresponding confidence level decreases and the confidence contours decrease in size. The closer the end point of the linked vector for the sample data falls to the center of the confidence contours, the more likely it is that the sample is from the null hypothesis.

Also, for a given sample size N, note that the i^{th} angle which is dependent solely on the auxiliary distribution remains unchanged and is used for all of the i^{th} linked vectors. Also, the magnitude of the sample data linked vector is solely dependent on the sample data set.

An example of the goodness-of-fit test is shown in Figure 2.4. The confidence contours are plotted for confidence contours of 0.9, 0.95, and 0.99. As mentioned above, if the end point of the sample data linked vector locus falls within a contour, then the sample data set is said to be statistically consistent with the null hypothesis at a confidence level based on the probability specified for that contour. If the sample data set is truly consistent with the null hypothesis, the system of sample linked vectors is likely to closely follow that for the system of null linked vectors.

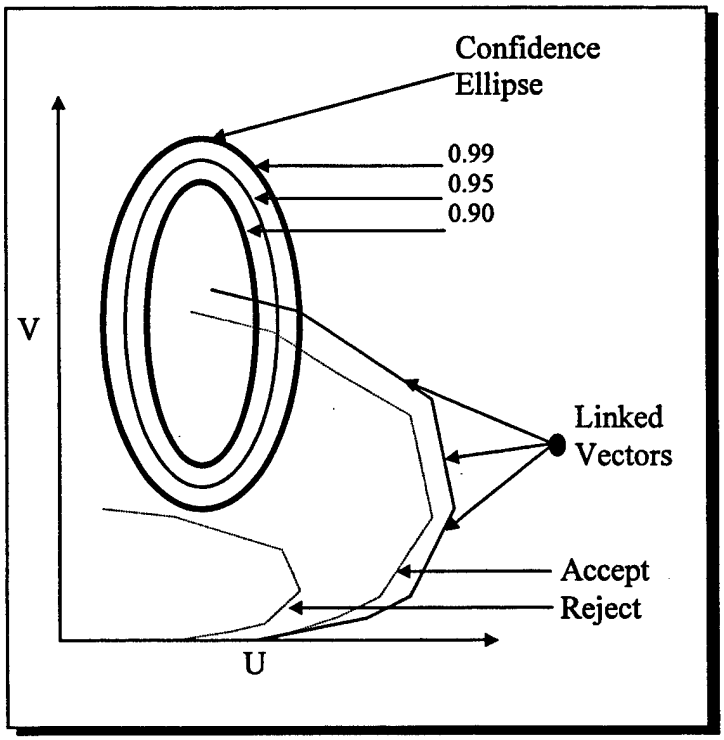


Figure 2.4 - Example of Goodness-of-fit Test

2.3.2.d - PDF Approximation Mode

The PDF approximation chart takes this a step further by providing other distributions. These distributions are computed in the same manner as described previously for the null hypothesis in that the magnitude of the linked vector is computed from the expected value of the ordered statistic of 2000 Monte-Carlo simulations. However, the angles θ_i are still computed from the auxiliary distribution, and, the confidence contours are computed only for the null hypothesis.

In order not to clutter the approximation chart, only the end points of all linked vectors are provided in the approximation chart, along with the confidence contours for the selected distribution (null hypothesis).

For distributions dependent only on mean and variance (no shape parameters), such as Gaussian, Uniform, and Cauchy, there exists only one unique linked vector (for a given value of N) since the data are normalized to have zero mean and unit variance. Thus, only one point on the approximation chart is plotted. For distributions dependent on a single shape parameter, such as Weibull, Lognormal, and K-distributed, different values of the shape parameter result in different linked vectors. Consequently, the end point of the linked vectors is also dependent on the shape parameter. The end points corresponding to different shape parameter values are joined to obtain a single curve on the approximation chart. This curve provides a unique representation for the PDF dependent on a single shape parameter for a given value of N. Similarly, for a distribution dependent on two shape parameters, such as the Beta distribution and SU-Johnson, a series of linked vectors must be computed in order to plot the surface on which the end point moves for varying shape parameters. This is formed by holding the first shape parameter constant and varying the second shape parameter to generate a curve, then changing the first shape parameter and again holding it constant while varying the second shape parameter, etc. until a family of curves is produced over the surface that the distribution occupies (for a given value of N).

The approximation chart is used by plotting on the chart the end point of the sampled linked vector $y_{i:N}$. To select the 'best' approximate PDF, the algorithm chooses the closest distribution to the end point and estimates the shape parameters of this

distribution if required. The algorithm can also provide a rank order of selection for PDF approximation of all available distributions based on their respective distances from the sample data.

Figure 2.5 shows an example of the approximation chart. Note that every point in the approximation chart corresponds to a specific distribution. That point closest to the sample data locus end point is chosen as the best approximation to the PDF underlying the random data. This closest point is determined by projecting the sample locus end point to all points on the approximation chart and selecting that point whose perpendicular distance from the sample point is the smallest. Once the PDF underlying the sample data is selected, the shape, location and scale parameters are then approximated.

2.3.3 - Details of the statistical procedure

Recall that the Ozturk algorithm not only approximates the body of the PDF for a given set of data but also gives confidence ellipses to determine the quality of the approximation. One can choose different PDFs such that their end points, defined by $[u_{100},v_{100}]$, and their corresponding 0.99 ellipses are spread throughout the approximation chart. By doing so, one can define regions in the Approximation Chart bounded by the ellipses where each region hosts the end-points of the data that would be generated from the corresponding PDF. In this case, whenever the end point of a given set of sample data falls within a given ellipse it will be assigned the PDF corresponding to that ellipse as the approximating PDF. This technique is detailed next.

Set-up of the Approximation Chart

Consider a set of 15 PDFs, as tabulated and numbered in Table 2.1 along with their shape parameters. The end points and the 99% confidence ellipses corresponding to the PDFs are displayed in the UV plane of the approximation chart, as shown in Figures 2.6 and 2.7, respectively. Figure 2.8 shows the colors assigned to the ellipses of the different PDFs as listed in Table 2.1

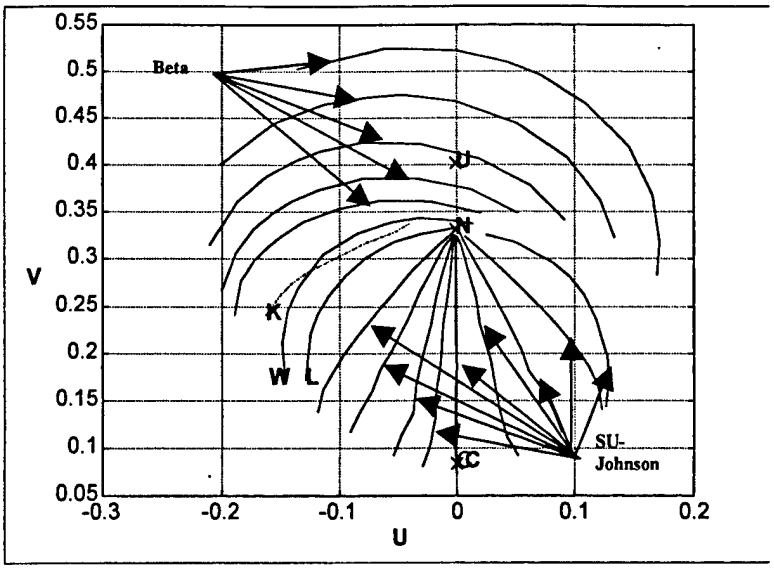


Figure 2.5 - Example of the Approximation Chart (U:Uniform, C:Cauchy, N:Normal, K:K-distributed, L:Lognormal, W:Weibull)

Color	16	15	14	13	12	11	10	9
No	2	1	74	70	29	101	97	38
PDF	U	N	SU	SU	B	SU	SU	SU
Shape 1	-	-	1.0	0.5	0.4	1.0	0.5	1.0
Shape 2	-	-	0.0	0.0	3.2	0.8	0.8	-0.7

Color	8	7	6	5	4	3	2	1
No	34	3	5	31	28	25	21	Out
PDF	SU	W	W	B	B	B	B	Out
Shape 1	0.5	0.6	1.1	1.0	6.0	6.0	6.0	-
Shape 2	-0.7	-	-	3.2	1.6	0.8	0.4	-

Table 2.1 - List of PDFs whose ellipses are used in the Approximation Chart and their corresponding colors.

N: Normal, W: Weibull, B: Beta, SU: SU-Johnson

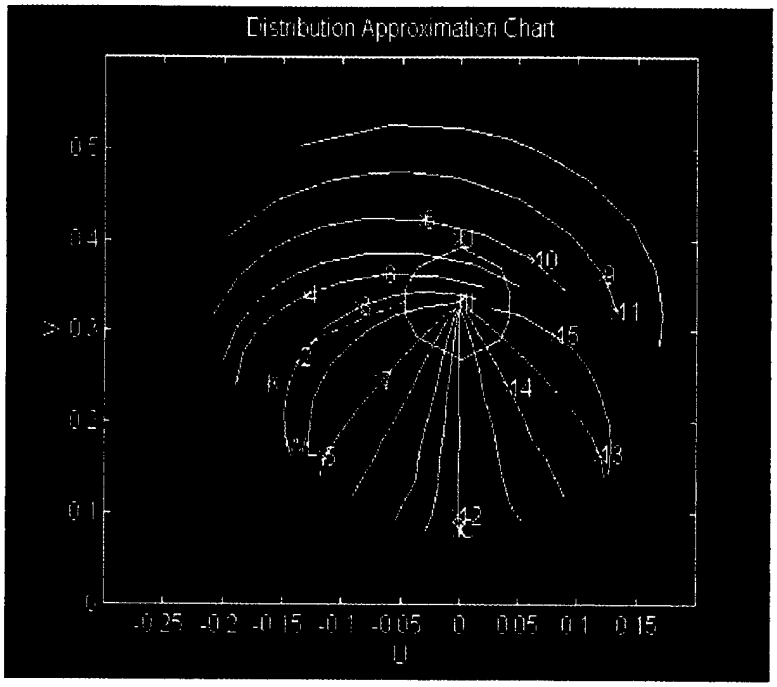


Figure 2.6 – End-Points location of the set of 15 PDFs

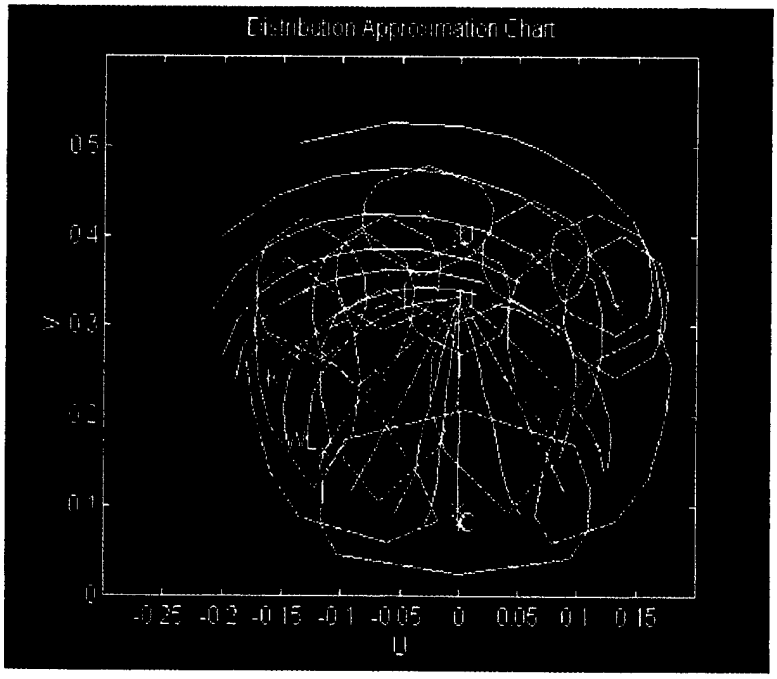


Figure 2.7 - 99% Ellipses of the set of 15 PDFs

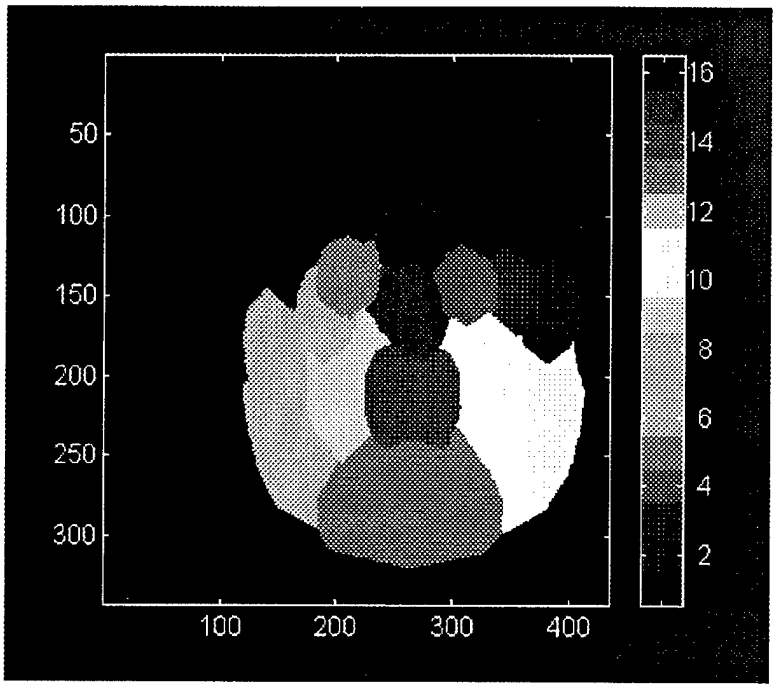


Figure 2.8 - Chart for the ellipses of the different PDFs

Set-up of the scene

✓ Consider now a scene consisting of MxN pixels. Let the pair (i,j) designate the location of a given pixel in the scene referred to as the test pixel. Each pixel (i,j) in the scene has a data value of v(i,j).

✓ For each test pixel select a set of 99 reference pixels which are the closest in location to the test pixel. Let the set of 100 data from the reference pixels and the test pixel be referred to as the sample data set for the test pixel. Note that in this work, 100 data samples are used each time to determine the best approximating PDF of the samples using the Ozturk algorithm.

Methodology

✓ For a given test pixel obtain its corresponding sample data set.

✓ Compute the coordinate $Q_{100}=[u_{100},v_{100}]$, as defined by Eqs. 6-8 corresponding to the data set.

✓ Check in which ellipse Q_{100} located.

✓ Record the number from Table 1 assigned to the PDF corresponding to that ellipse.

✓ Assign that number to the test pixel.

✓ Repeat the above steps for each pixel in the scene.

At the end, a mapped scene is obtained where each pixel would have a value between 1 and 15 referring to the PDF which would best approximate the data in the test pixel when it is associated with its closest 99 reference pixels. The mapped scene will show "regions" with different values. These regions will define the homogeneous patches in the scene.

Chapter III

Application of A'SCAPE to Medical Imaging

3.1 - Introduction

The A'SCAPE procedure is applied next to real data of CT scans. First, one example of a known benign case and a second example of a known malignant case are used to build rules to identify the tumors and classify them. Then, the rules are tested by applying them to three different cases of unknown types of tumor. Note that, because only one sample from each case (malignant and benign) is used to build the rules, these are not necessary well justified and may be erroneous. On the other hand, these rules serve the objective of showing how to approach the problem of detecting and classifying tumors using the A'SCAPE procedure.

The Mapping procedure is used to isolate all regions that might be identified as tumors. This is done by clearing all non-desirable patches such as branches and vessels from the scene. Among the patches left in the scene, those with more then 100 pixels are then identified to be processed later by the Statistical procedure. The latter procedure is used to both identify and classify tumors.

In the following sections, examples of the application of A'SCAPE are shown. First in Sections 3.2 and 3.3, A'SCAPE is applied to known cases of benign and malignant tumors. Then, in Section 3.4, rules are built to help identify and classify tumors. Finally, in Sections 3.5-3.7, the rules are tested on examples of cases of lungs with unknown types of tumor.

3.2 – Example 1: Benign Case Number E20799S2I1

Consider the case of a lung with a Benign tumor shown in the lower lung of Figure 3.1. Note that the original scene is a 512 by 512 image. First, the lung with the tumor is selected as shown in Figure 3.2. Note that this scene has 130 by 149 pixels. The tumor is shown by the patch inside the lung located in the bottom left of the image. Application of the Mapping procedure results in 5 thresholds generating, thus, 5

component images whose composite image is shown in Figure 3.3 with 5 different gray levels. The component image that contains the tumor is selected as shown in Figure 3.4. The component image is then labeled to result in the labeled scene of Figure 3.5. Note that, as shown in Table 3.1, only patch number 3 meets the condition that it is inside the lung and contains more than 100 pixels. This patch is in fact the tumor to be investigated. Table 3.2 summaries the type of approximating PDFs for each pixel in patch 3 as well as the number of pixels with the same approximating PDF.

Processing patch number 3 through the Statistical Procedure results in the colored scene of Figure 3.6. Note that the different colors in the scene refer to the PDFs of the corresponding pixels as listed in Table 2.1. Also, Figure 3.7 shows in black the spread of the UV pairs corresponding to the pixels of patch number 3. The following are to be noted:

❒ In Figure 3.6,

✓ two distinct contiguous regions can be seen: Region 1 formed by pixels with UV pairs outside the ellipses, and, Region 2 formed by pixels from the Beta PDF surrounding other pixels from the SU-Johnson PDF

✓ concentric behavior is displayed by the pixels from the Beta PDF surrounding other pixels from the SU-Johnson PDF

❒ In Figure 3.7,

✓ UV pairs are located on the right part of the chart

✓ UV pairs are wide spread

✓ UV pairs are spread vertically

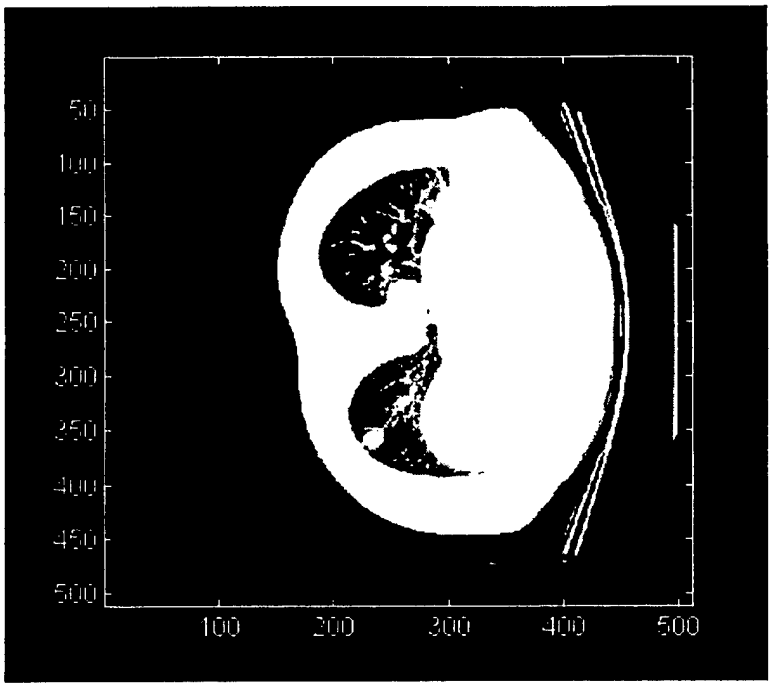


Figure 3.1 - E20799S2I1- Original Scene

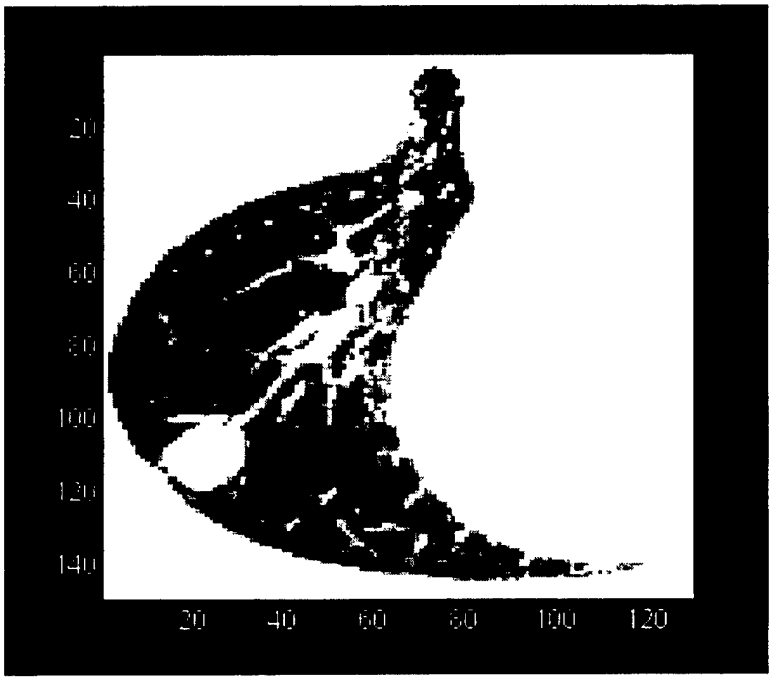


Figure 3.2 - E20799S2I1- Lung with Tumor

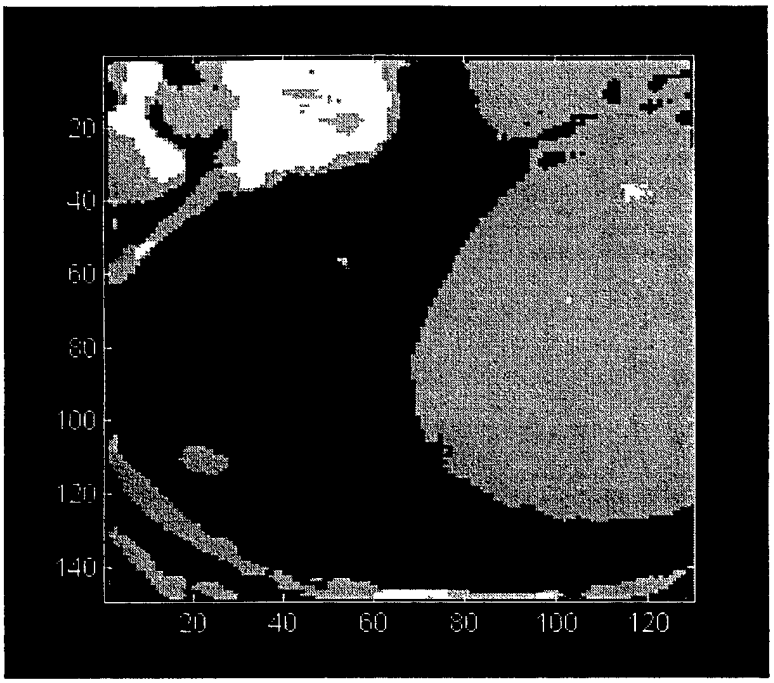


Figure 3.3 - E20799S2I1- Composite scene

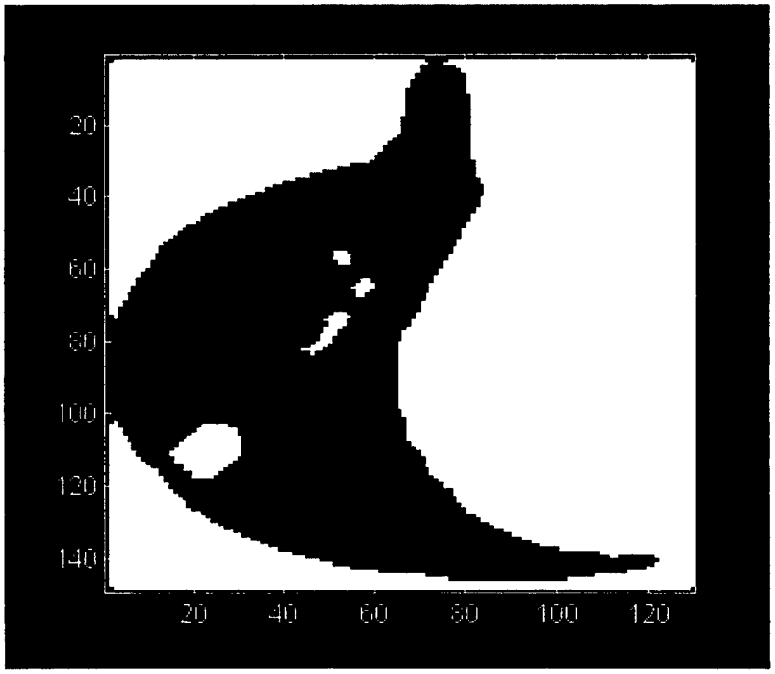


Figure 3.4 - E20799S2I1- Component Scene Containing Tumor

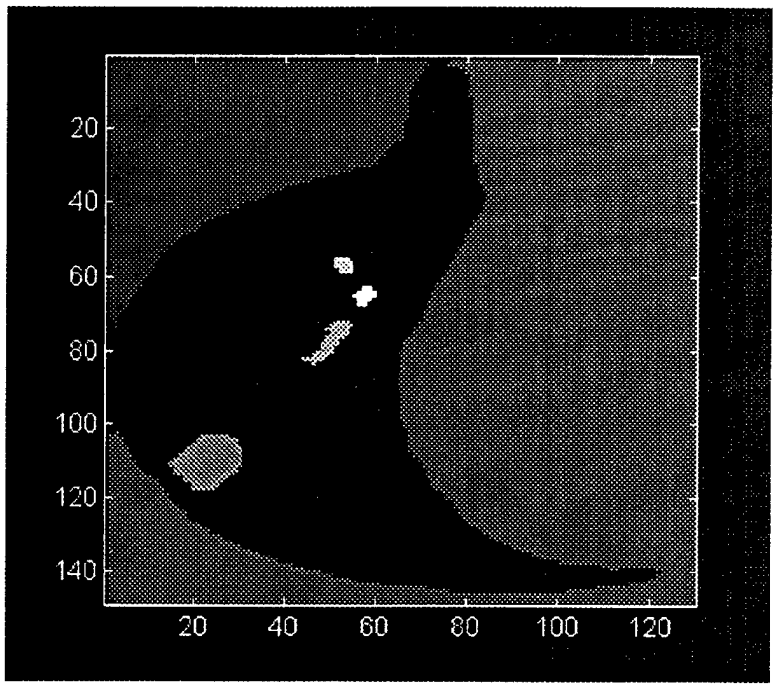


Figure 3.5 - E20799S2I1- Labeled Scene

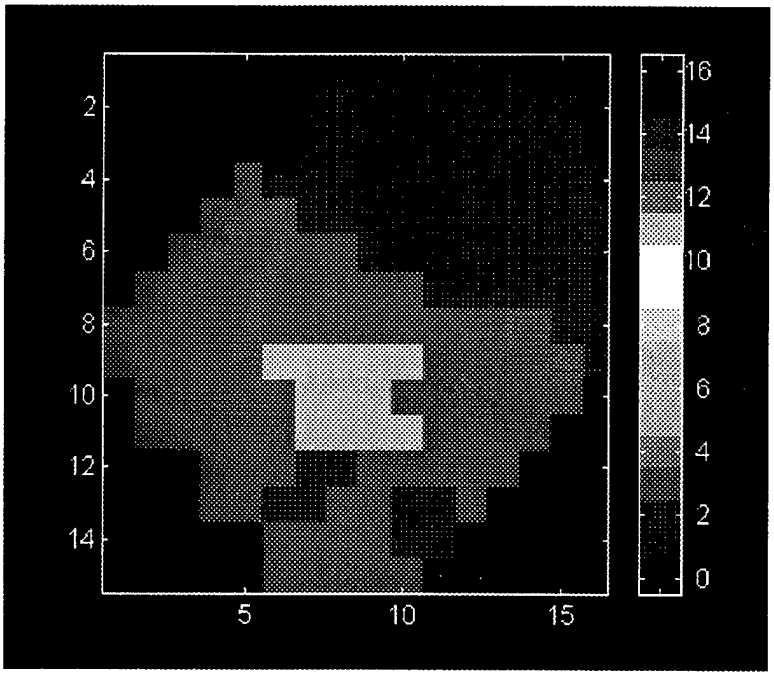


Figure 3.6 - E20799S2I1 - Patch No. 3 with Approximated PDFs

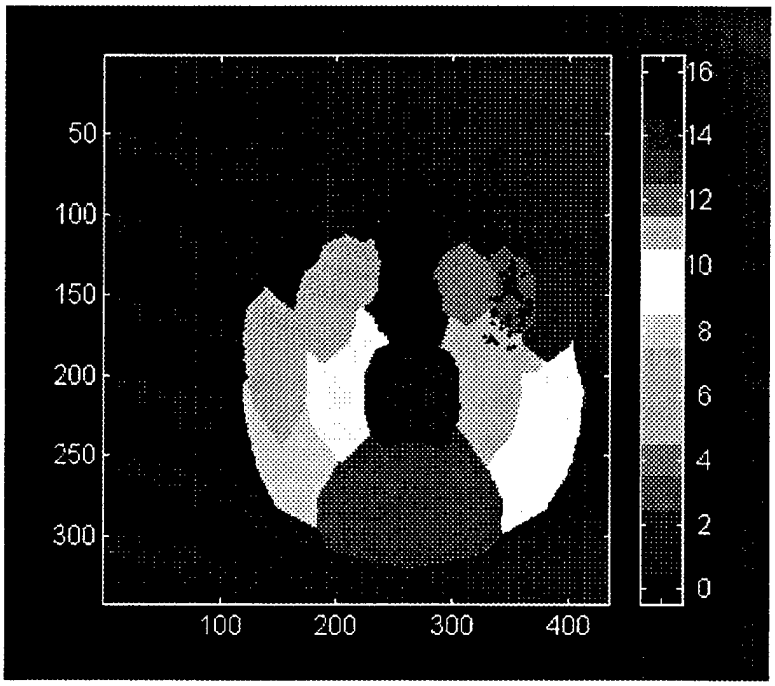


Figure 3.7 - E20799S2I1- Location of UV pairs for Patch No. 3

Patch No.	1	2	3	4	5	6
No. of Pixels	2636	8884	168	45	14	16
Location wrt Lung	Out	Out	In	In	In	In

Table 3.1 - E20799S2I1 – Results of labeling

Patch No.	3			
PDF No.	1	2	3	11
No. of Pixels	63	8	85	12

Table 3.2 - E20799S2I1 – PDFs of the pixels in patch 3

3.3 – Example 2: Malignant Case Number E14649S2I14

Consider now the case of a lung with a Malignant tumor shown in the upper lung of Figure 3.8. Note that the original scene is a 512 by 512 image. As was done in example 1, the lung with the tumor is selected as shown in Figure 3.9. Note that this scene has 278 by 191 pixels. The tumor is shown by the patch inside the lung located in the upper left of the image. Application of the Mapping procedure results in 6 thresholds generating, thus, 6 component images whose composite image is shown in Figure 3.10 with 6 different gray levels. The component image that contains the tumor is selected as shown in Figure 3.11. The component image is then labeled to result in the labeled scene of Figure 3.12. Note that, as shown in Table 3.3, only patch numbered 2 is inside the lung and has more than 100 pixels. This patch is in fact the tumor to be investigated. Table 3.4 summaries the type of approximating PDFs for each pixel in patch 2 as well as the number of pixels with the same approximating PDF.

Processing patch numbered 2 through the Statistical Procedure results in the colored scene of Figure 3.13. Note again that the different colors in the scene refer to the PDFs of the corresponding pixels as listed in Table 2.1. Also, Figure 3.14 shows in black the spread of the UV pairs corresponding to the pixels of patch numbered 2. The following are to be noted:

❒ In Figure 3.13,

✓ two distinct contiguous regions can be seen: Region 1 formed by pixels with UV pairs in a particular Beta's ellipse, and, Region 2 formed by pixels from another Beta PDF

❒ In Figure 3.14,

✓ UV pairs are located on the right part of the chart

✓ UV pairs are wide spread

✓ UV pairs are spread horizontally

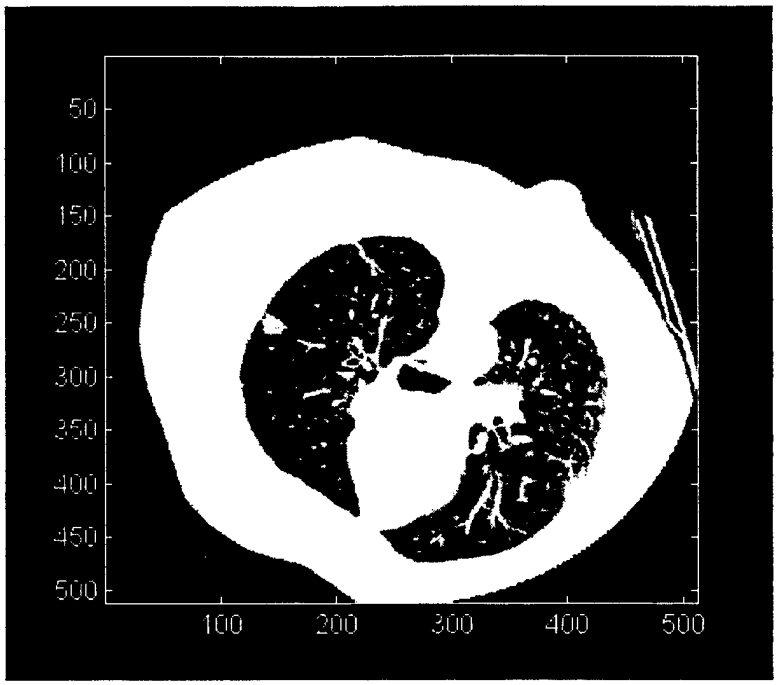


Figure 3.8 - E14649S2I14 - Original Scene

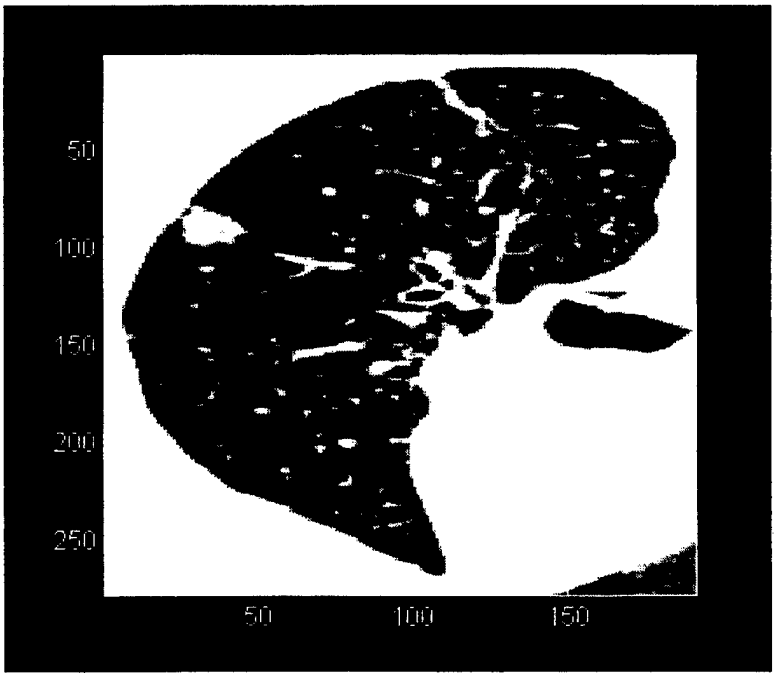


Figure 3.9 - E14649S2I14 - Lung with Tumor

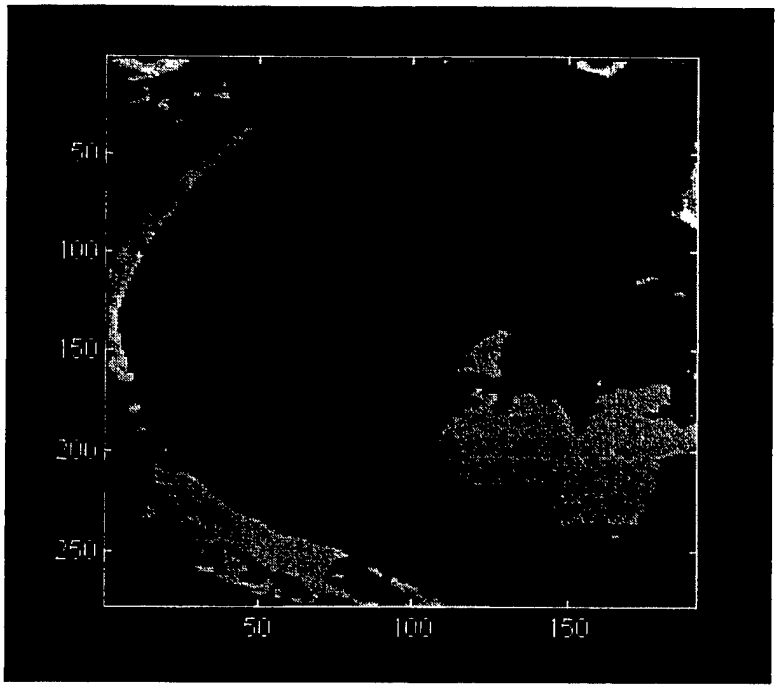


Figure 3.10 - E14649S2I14 - Composite Scene

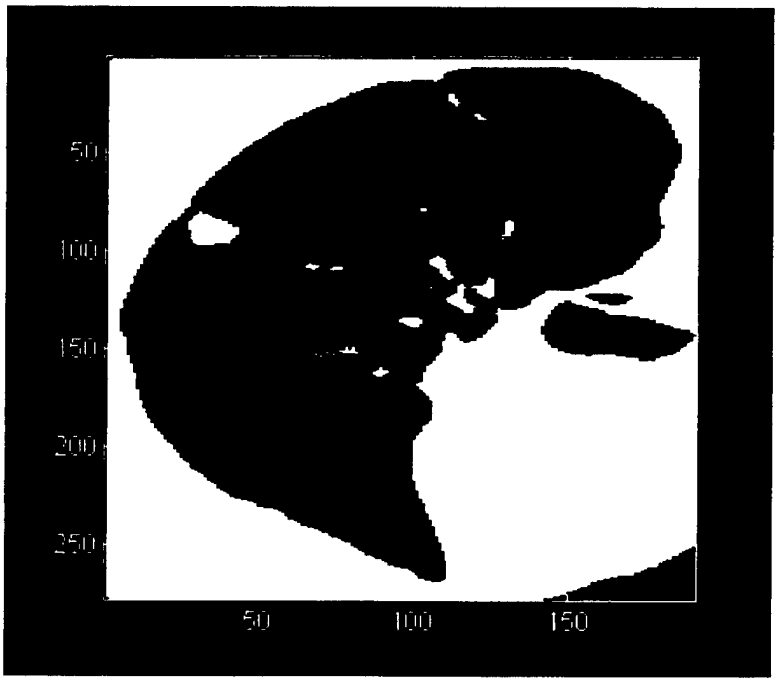


Figure 3.11 - E14649S2I14 - Component Scene Containing Tumor

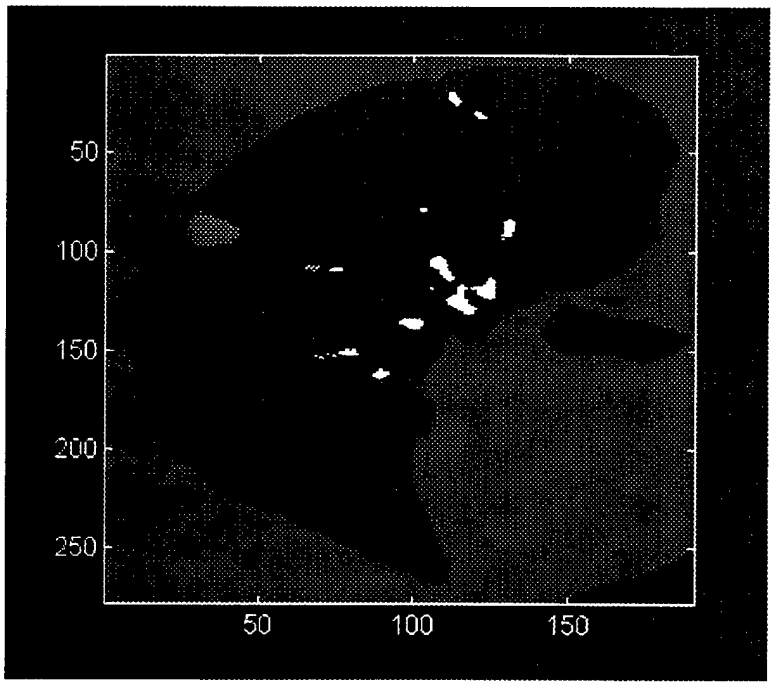


Figure 3.12 - E14649S2I14 - Labeled Scene

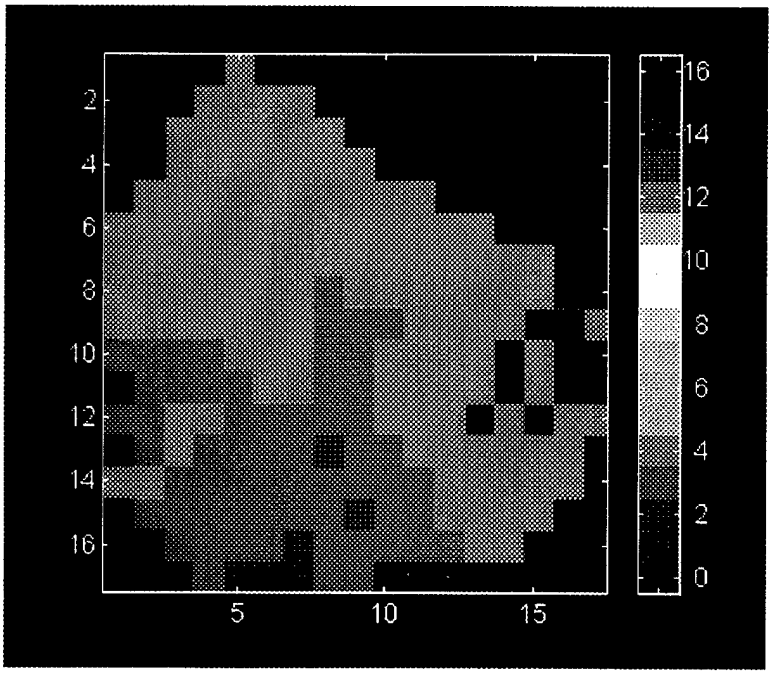


Figure 3.13 - E14649S2I14 - Patch No. 2 with Approximated PDFs

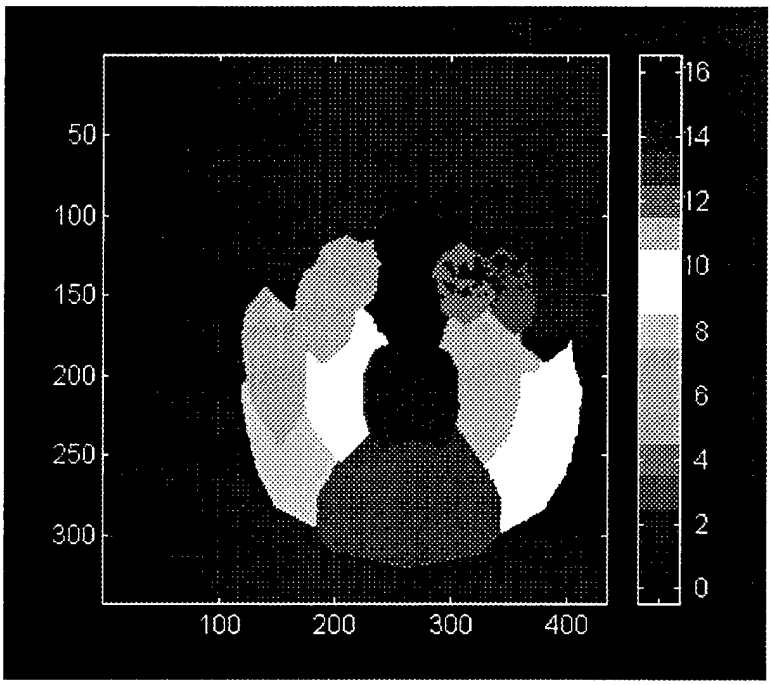


Figure 3.14 - E14649S2I14 - Location of UV pairs for Patch No. 2

Patch No.	1	2	3	4	5	6	7	8	9	10
No. of Pixels	24168	201	11	1	1	1	7	1	12	16
Location wrt Lung	Out	In	In	In	In	In	In	In	In	In
Patch No.	11	12	13	14	15	16	17	18	19	20
No. of Pixels	27	4	46	1	60	14	47	7	21	1
Location wrt Lung	In	In	In	In	In	In	In	In	In	In

Table 3.3 - E14649S2I14 - Results of labeling

Patch No.	2				
PDF No.	1	2	3	4	15
No. of Pixels	4	2	60	125	10

Table 3.4 - E14649S2I14 - PDFs of the pixels in patch 3

3.4 – Rules deducted from examples 1 and 2

Using the notes on Figures 3.6, 3.7, 3.13, 3.14, on the Benign and Malignant tumors, the following rules are deducted:

Rule number 1: A patch is declared to be a Tumor when

- ❒ UV pairs are located on the right part of the chart
- ❒ UV pairs are spread

Rule number 2: A patch is declared to be a Benign Tumor when

- ❒ UV pairs are spread vertically
- ❒ Concentric regions are displayed in the colored patch

Rule number 3: A patch is declared to be a Malignant Tumor when

- ❒ UV pairs are spread horizontally
- ❒ Contiguous regions are displayed in the colored patch

It is important to mention again that the number of samples upon which the rules are based is very limited and, thus, the rules might not be valid. Using the data at hand and the above rules, three cases of unknown types of tumor are investigated next. In each case, even-though the tumor location is known a-priori, its type (Malignant or Benign) is unknown. Note that for each case (1) the Mapping procedure segments the scene, cleans out branches, and detects patches that might be tumors. Then, (2) for each patch with at least 100 pixels, the Statistical procedure finds the PDFs that can approximate the statistical distribution for the different pixels, and, assigns different colors for pixels with different approximating PDFs. Finally, (3) the above rules are applied to each patch and decisions are made to answer the following questions:

(1) is the patch a tumor,

(2) if the patch is a tumor, is it Malignant or Benign.

3.5 – Example 3: Unknown Case Number E7066S2I17

Consider first the case of a lung with a tumor whose type is unknown. The tumor is located in the lower lung of Figure 3.15. Note that the original scene is a 512 by 512 image. As was done in examples 1 and 2, the lung with the tumor is selected as shown in Figure 3.16. Note that this scene has 218 by 284 pixels. The tumor is shown by the patch inside the lung located in the lower right of the image. Application of the Mapping procedure results in the composite image of Figure 3.17. The component image that contains the tumor is selected as shown in Figure 3.18. The patches in the component image are then labeled to result in the scene of Figure 3.19. Note that, as shown in Table 3.5, patches numbered 3, 4, and 10 are inside the lung and have more than 100 pixels each. Patch numbered 10 is the tumor. For each patch, Table 3.6 summaries the type of approximating PDFs for each pixel as well as the number of pixels with the same approximating PDF.

Processing patch numbered 3 through the Statistical Procedure results in the colored scene of Figure 3.20. Also, Figure 3.21 shows in black the spread of the UV pairs corresponding to the pixels of patch numbered 3. The following are to be noted:

- In Figure 3.20,
  - ✓ most of the pixels have the same color (e.g., PDF), no contiguous or concentric regions are observed
- In Figure 3.21,
  - ✓ UV pairs are located on the left part of the chart
  - ✓ UV pairs are not wide spread

With respect to rules 1 to 3 of Section 3.3, it is noted that (1) the UV pairs are located on the left part of the chart and not on the right part, and that they are not spread, and, (2) the colored patch does not display any contiguous or concentric behavior of colors. Thus, it is concluded that **the patch is not a tumor.**

Processing patch numbered 4 through the Statistical Procedure results in the colored scene of Figure 3.22. Also, Figure 3.23 shows in black the spread of the UV pairs corresponding to the pixels of patch numbered 4. The following are to be noted:

❒ In Figure 3.22,

✓ all pixels have the same PDF, no contiguous or concentric regions are observed

❒ In Figure 3.23,

✓ UV pairs are located on the middle part of the chart

✓ UV pairs are not wide spread

With respect to rules 1 to 3 of Section 3.3, it is noted that (1) the UV pairs are located on the middle part of the chart and not on the right part, and that they are not spread, and, (2) the colored patch does not display any contiguous or concentric behavior of colors. Thus, it is concluded that **the patch is not a tumor**.

Processing patch numbered 10 through the Statistical Procedure results in the colored scene of Figure 3.24. Also, Figure 3.25 shows in black the spread of the UV pairs corresponding to the pixels of patch numbered 10. The following are to be noted:

❒ In Figure 3.24,

✓ concentric regions are observed in the colored patch

❒ In Figure 3.25,

✓ UV pairs are located on the right part of the chart

✓ UV pairs are wide spread

With respect to rules 1 to 3 of Section 3.3, it is noted that (1) because the UV pairs are located on the right part of the chart and that they are wide spread, it is concluded **that the patch is a tumor**. Also, (2) because the colored patch displays a concentric behavior of colors, it is concluded that **the patch is a Benign tumor**.

Note that in this example all patches were correctly identified with respect to containing a tumor.

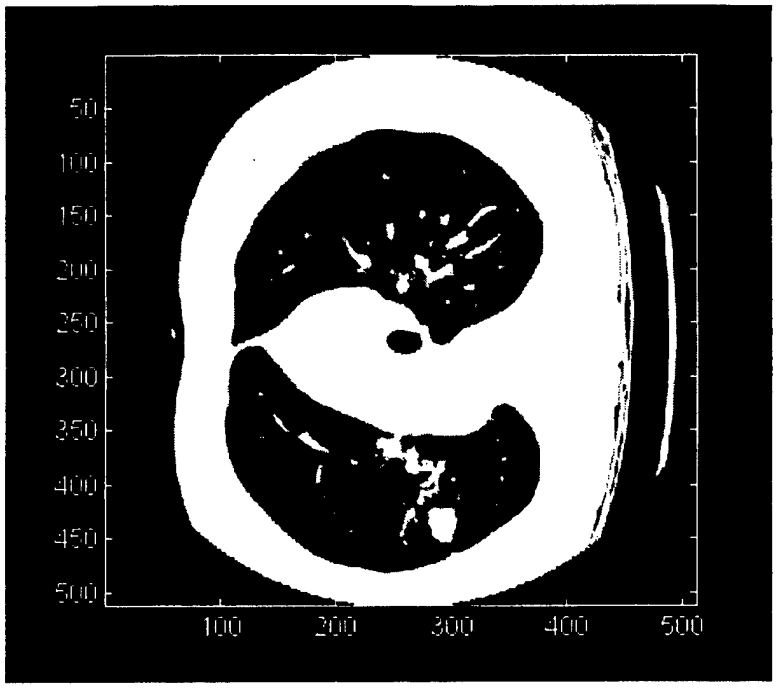


Figure 3.15 - E7066S2I17 - Original Scene

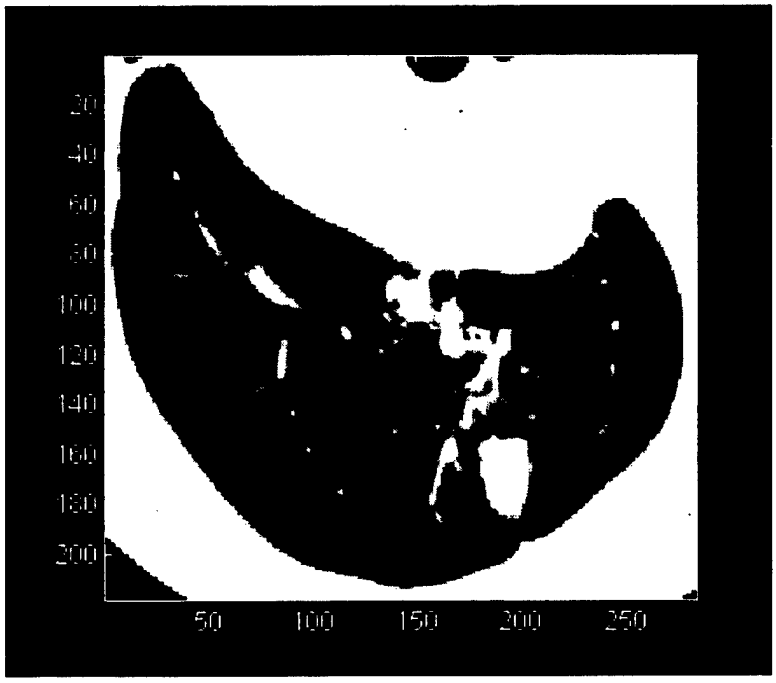


Figure 3.16 - E7066S2I17 - Lung with Tumor

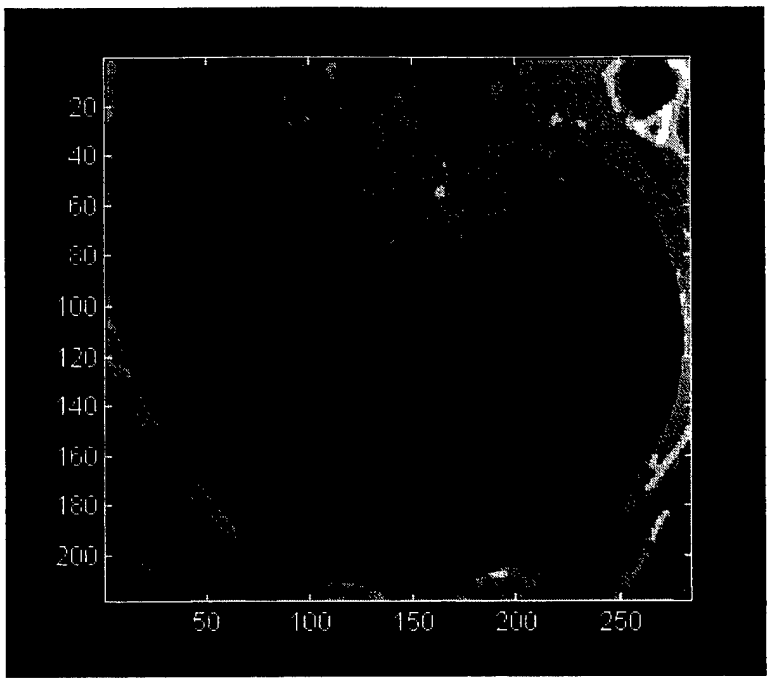


Figure 3.17 - E7066S2I17 - Composite scene

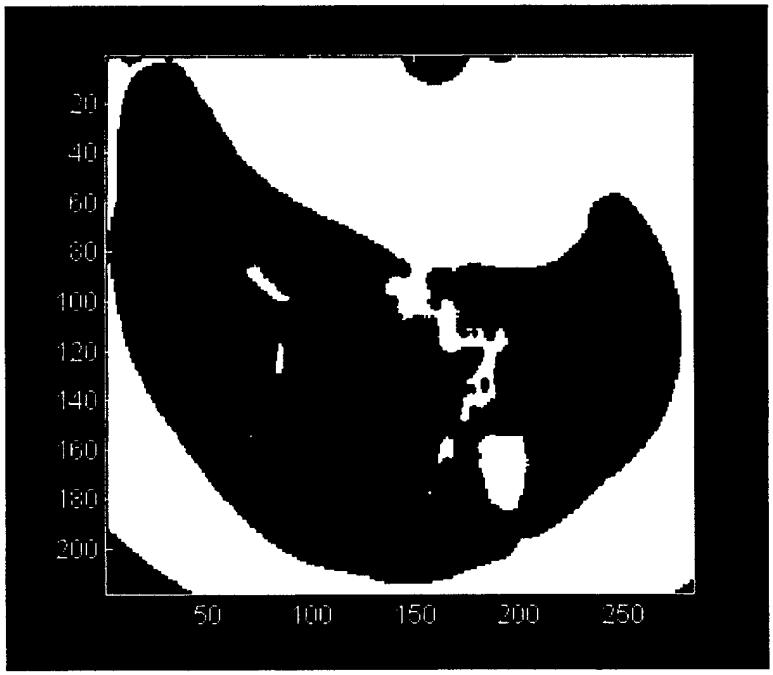


Figure 3.18 - E7066S2I17 - Component Scene Containing Tumor

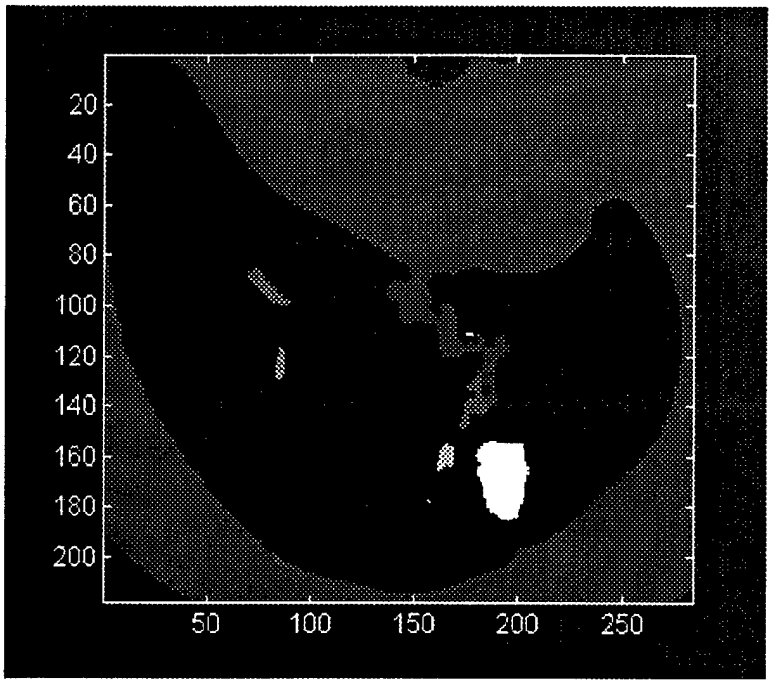


Figure 3.19 - E7066S2I17 - Labeled Scene

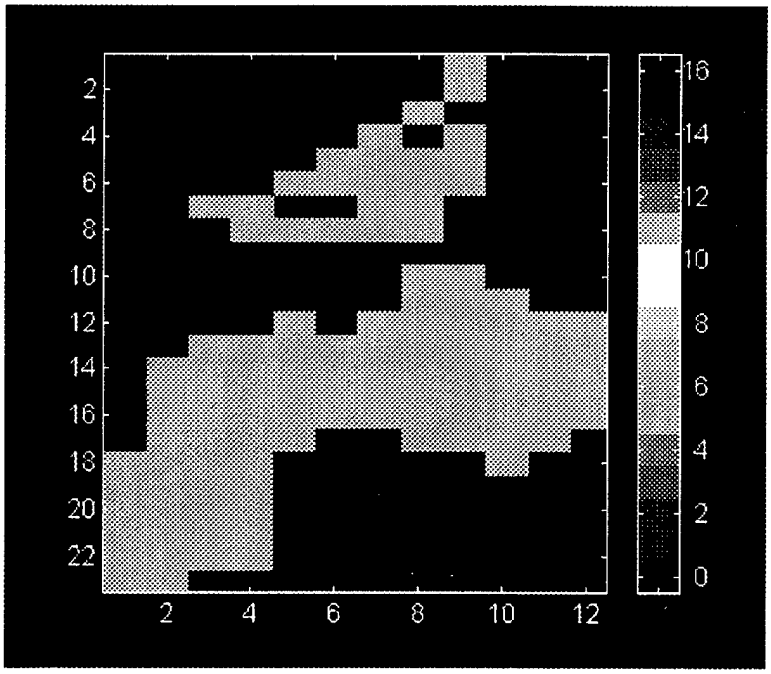


Figure 3.20 - E7066S2I17 - Patch No. 3 with Approximated PDFs

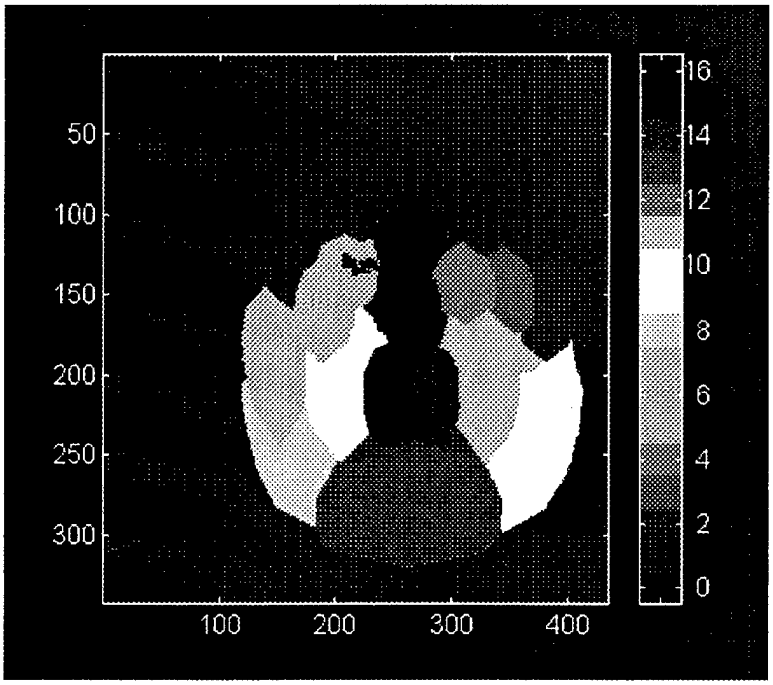


Figure 3.21 - E7066S2I17 - Location of UV pairs for Patch No. 3

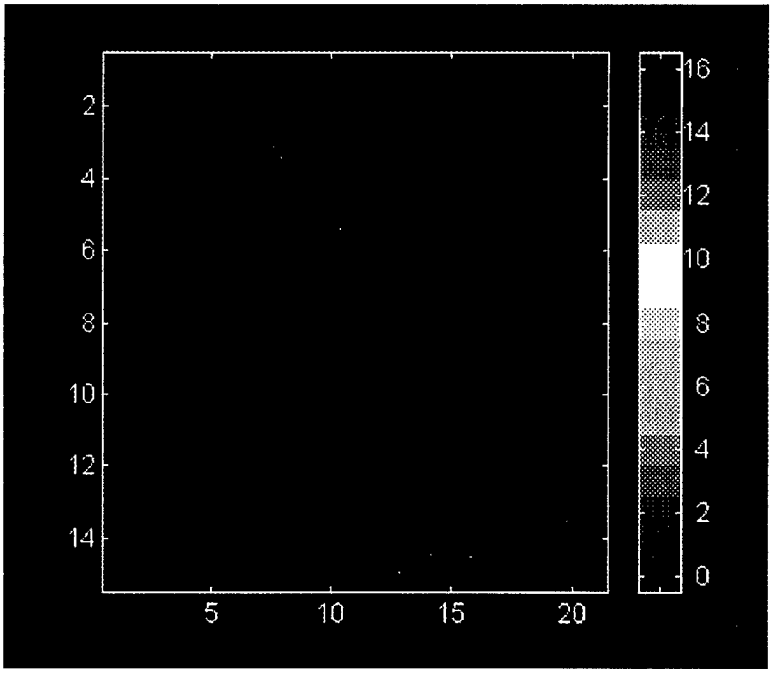


Figure 3.22 - E7066S2I17 - Patch No. 4 with Approximated PDFs

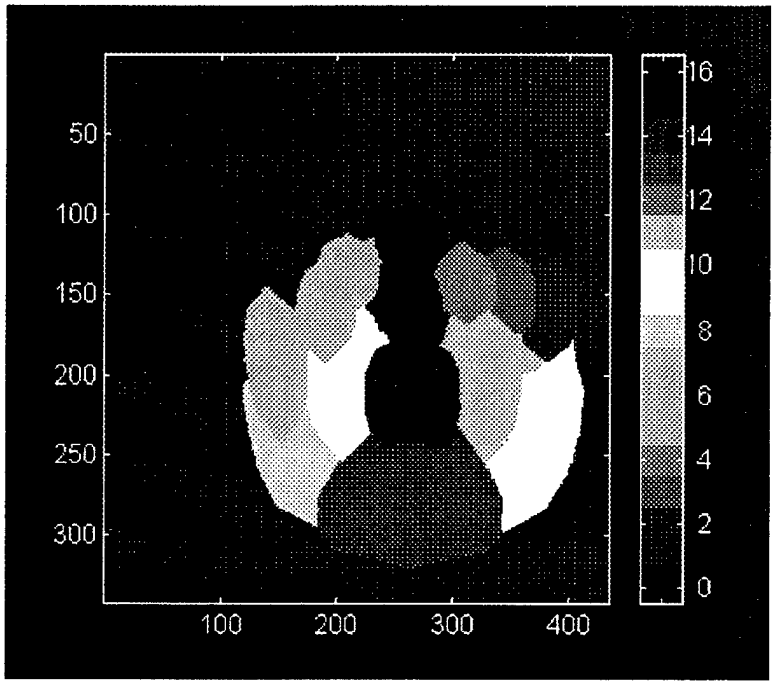


Figure 3.23 - E7066S2I17 - Location of UV pairs for Patch No. 4

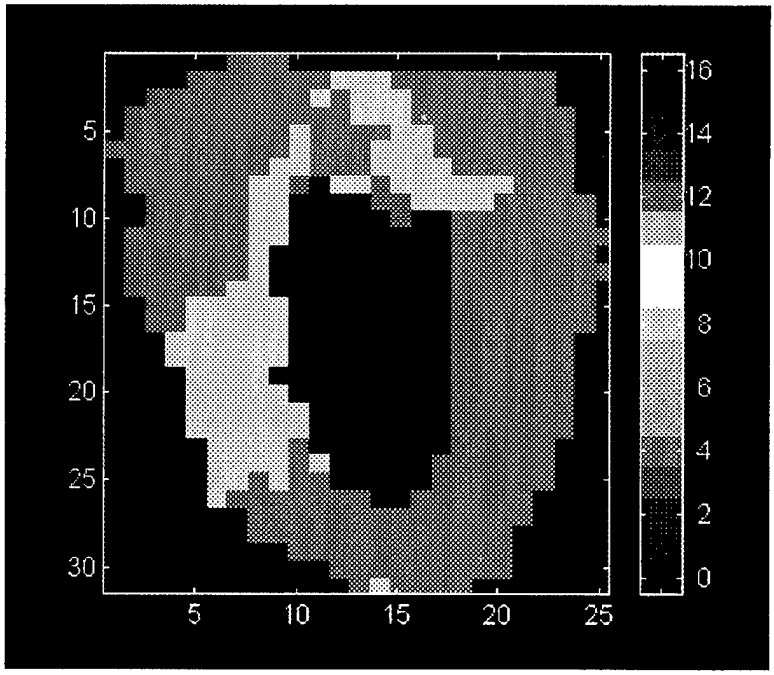


Figure 3.24 - E7066S2I17 - Patch No. 10 with Approximated PDFs

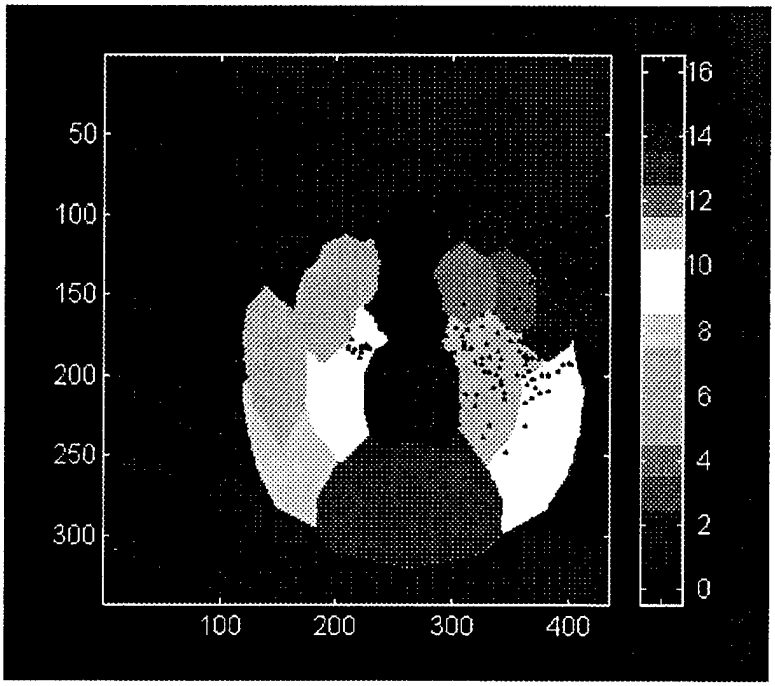


Figure 3.25 - E7066S2I17 - Location of UV pairs for Patch No. 10

Patch No.	1	2	3	4	5	6	7	8	9	10
No. of Pixels	486	26707	120	102	35	1	59	2	3	570
Location wrt Lung	Out	In	In	In	In	In	In	In	In	In

Table 3.5 - E7066S2I17 - Results of labeling

Patch No.	3								
PDF No.	5	15							
No. of Pixels	109	11							
Patch No.	4								
PDF No.	16								
No. of Pixels	102								
Patch No.	10								
PDF No.	1	2	3	4	9	10	11	14	15
No. of Pixels	332	104	5	1	12	22	45	43	6

Table 3.6 - E7066S2I17 - PDFs of the pixels in patches 3, 4, and 10

3.6 – Example 4: Unknown Case Number E5424S2I15

Consider the case of a lung with a tumor whose type is unknown. The tumor is located in the lower lung of Figure 3.26. Note that the original scene is a 512 by 512 image. As was done in the previous examples, the lung with the tumor is first selected as shown in Figure 3.27. Note that the scene has been reduced to 197 by 264 pixels. The tumor is shown by the patch inside the lung located in the lower right of the image. Application of the Mapping procedure results in the composite image of Figure 3.28. The component image that contains the tumor is selected as shown in Figure 3.29. The patches in the component image are then labeled to result in the scene of Figure 3.30. Note that, as shown in Table 3.7, patches numbered 5, 7, and 13 are inside the lung and have more than 100 pixels each. Patch numbered 13 is the tumor. For each patch, Table 3.8 summaries the type of approximating PDFs for each pixel as well as the number of pixels with the same approximating PDF.

Processing patch numbered 5 through the Statistical Procedure results in the colored scene of Figure 3.31. Also, Figure 3.32 shows in black the spread of the UV pairs corresponding to the pixels of patch numbered 5. The following are to be noted:

- In Figure 3.31,
  - ✓ concentric regions are observed in the colored patch
- In Figure 3.32,
  - ✓ UV pairs are located on the right part of the chart
  - ✓ UV pairs are wide spread
  - ✓ UV pairs are spread vertically

With respect to rules 1 to 3 of Section 3.3, it is noted that (1) because the UV pairs are located on the right part of the chart and that they are wide spread, it is concluded **that the patch is a tumor**. Also, (2) because the colored patch displays a concentric behavior of colors, and that the UV pairs are spread vertically, it is concluded that **the patch is a Benign tumor**.

Processing patch numbered 7 through the Statistical Procedure results in the colored scene of Figure 3.33. Also, Figure 3.34 shows in black the spread of the UV pairs corresponding to the pixels of patch numbered 7. The following are to be noted:

❒ In Figure 3.33,

✓ most of the pixels have the same PDF, no contiguous or concentric regions are observed

❒ In Figure 3.34,

✓ UV pairs are located on the middle part of the chart

✓ UV pairs are not wide spread

With respect to rules 1 to 3 of Section 3.3, it is noted that (1) the UV pairs are located on the middle part of the chart and not on the right part, and that they are not spread, and, (2) the colored patch does not display any contiguous or concentric behavior of colors. Thus, it is concluded that **the patch is not a tumor**.

Processing patch numbered 13 through the Statistical Procedure results in the colored scene of Figure 3.35. Also, Figure 3.36 shows in black the spread of the UV pairs corresponding to the pixels of patch numbered 13. The following are to be noted:

❒ In Figure 3.35,

✓ concentric regions are observed in the colored patch

✓ contiguous regions are observed in the upper right part of the colored patch

❒ In Figure 3.36,

✓ UV pairs are located mostly on the right part of the chart

✓ UV pairs are wide spread

With respect to rules 1 to 3 of Section 3.3, it is noted that (1) because the UV pairs are located on the right part of the chart and that they are wide spread, it is concluded **that the patch is a tumor**. Also, (2) because the colored patch displays a concentric behavior of colors, it is concluded that **the patch is a Benign tumor**. In addition (3) the upper part

of the chart displays a contiguous behavior of colors, it is also concluded **that the upper part of the patch is a Malignant tumor**.

Note that in this example not all patches were correctly identified with respect to containing a tumor. Namely, patch 3 was identified as containing a tumor; in fact, it did not.

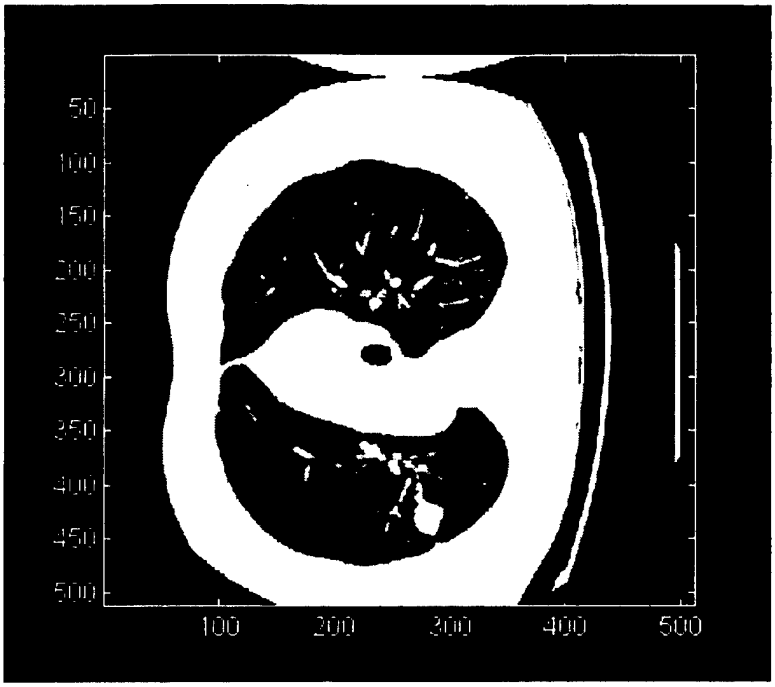


Figure 3.26 - E5424S2I15 - Original Scene

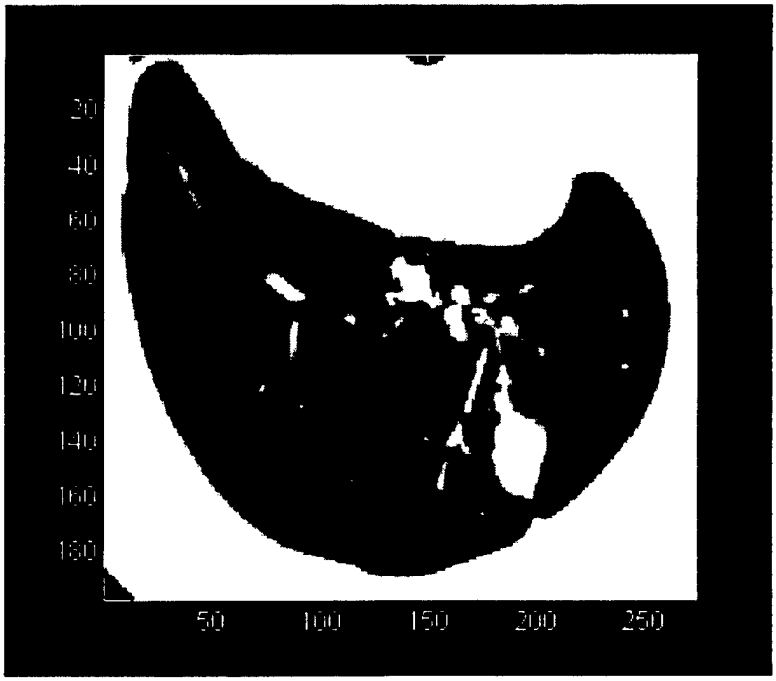


Figure 3.27 - E5424S2I15 - Lung with Tumor

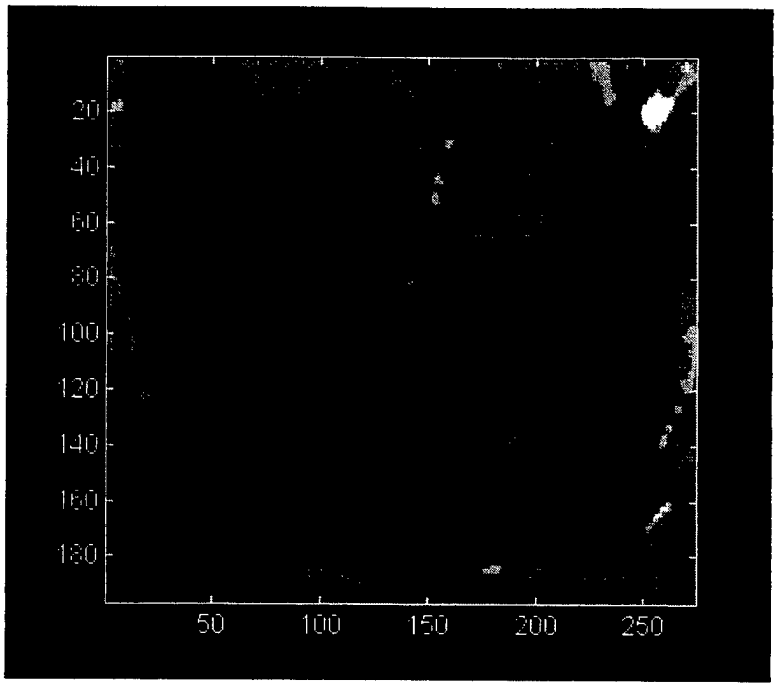


Figure 3.28 - E5424S2I15 - Composite scene

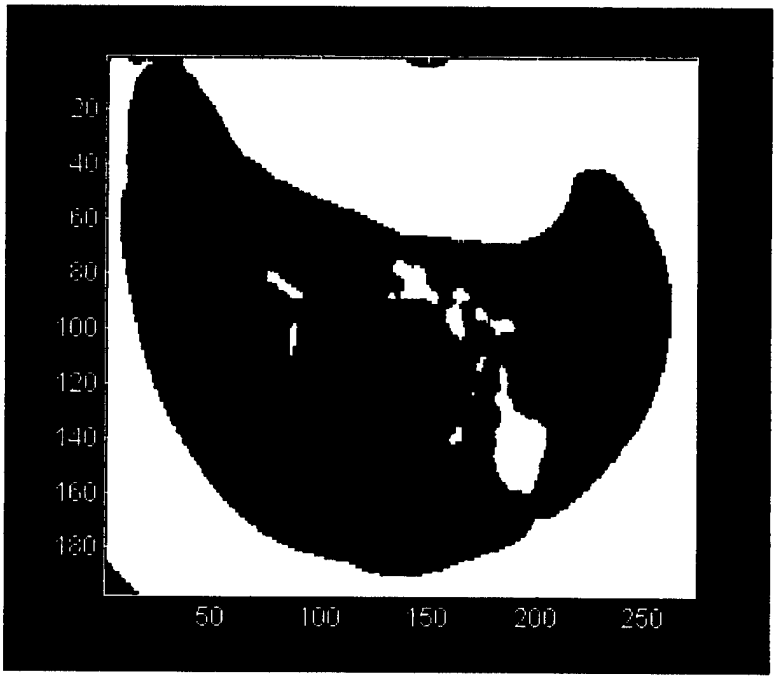


Figure 3.29 - E5424S2I15 - Component Scene Containing Tumor

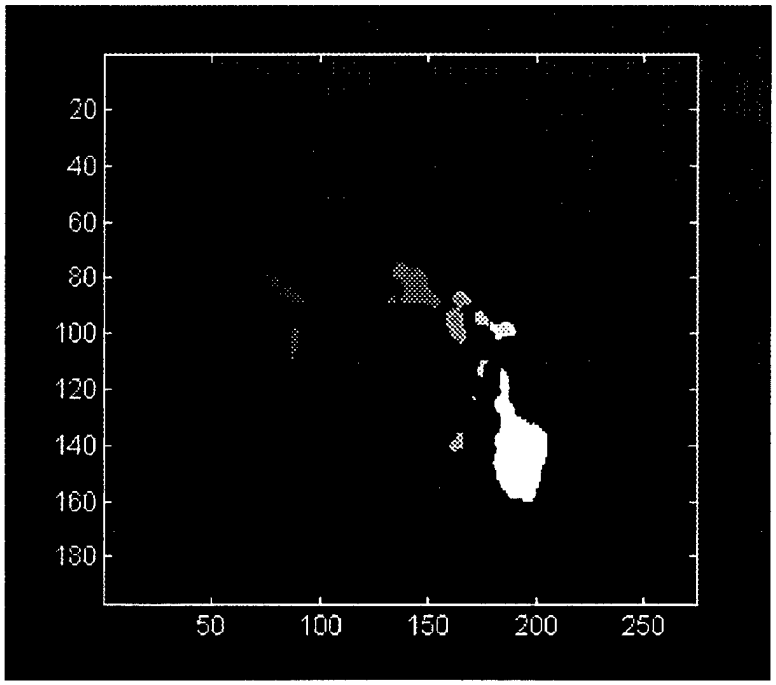


Figure 3.30 - E5424S2I15 - Labeled Scene

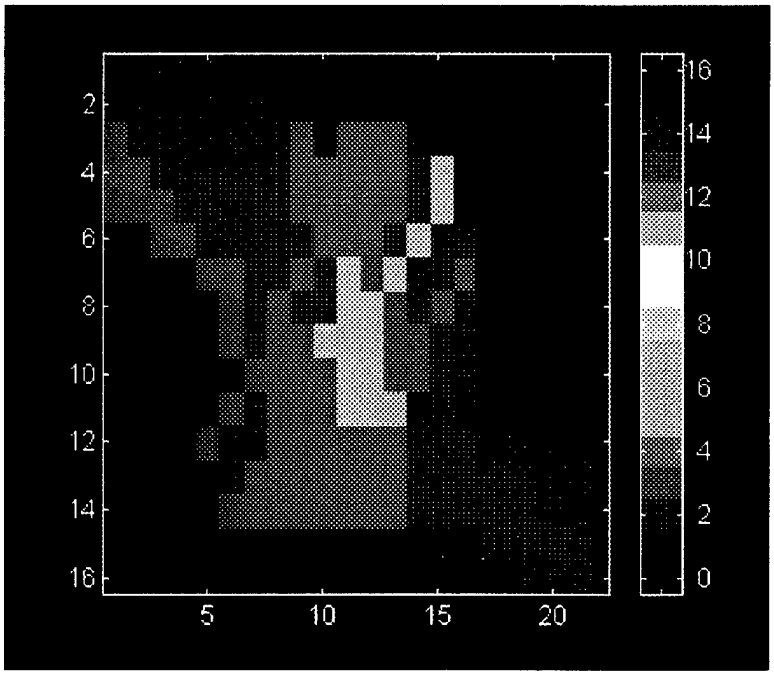


Figure 3.31 - E5424S2I15 - Location of UV pairs for Patch No. 5

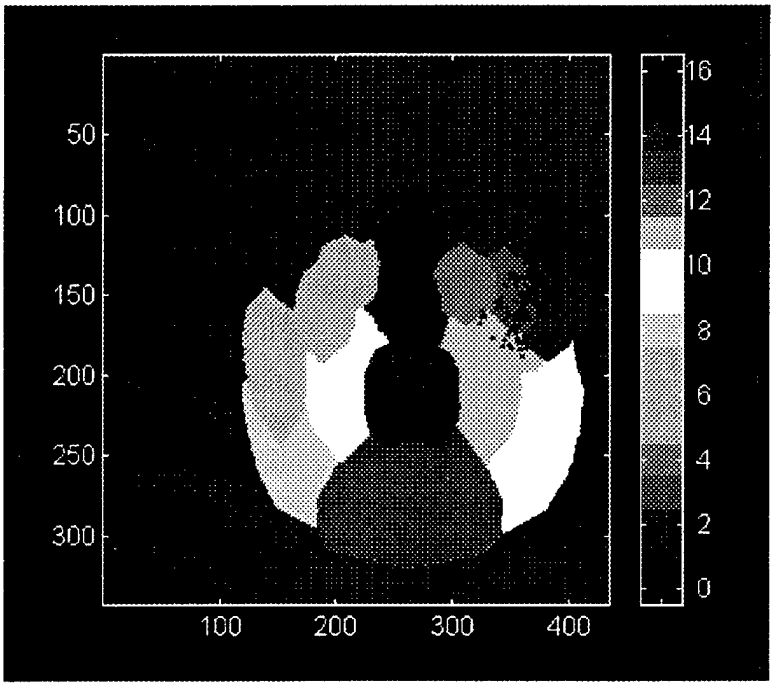


Figure 3.32 - E5424S2I15 - Location of UV pairs for Patch No. 5

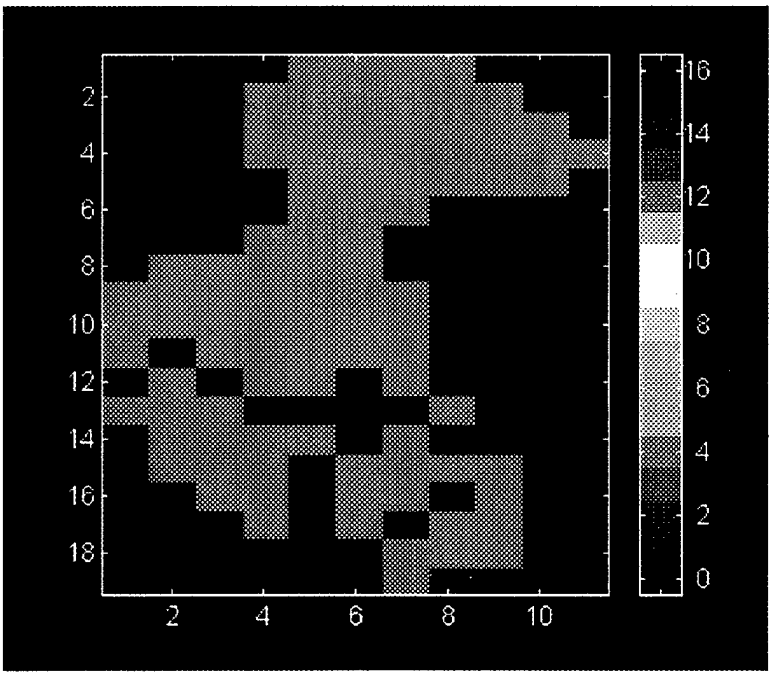


Figure 3..33 - E5424S2I15 - Patch No. 7 with Approximated PDFs

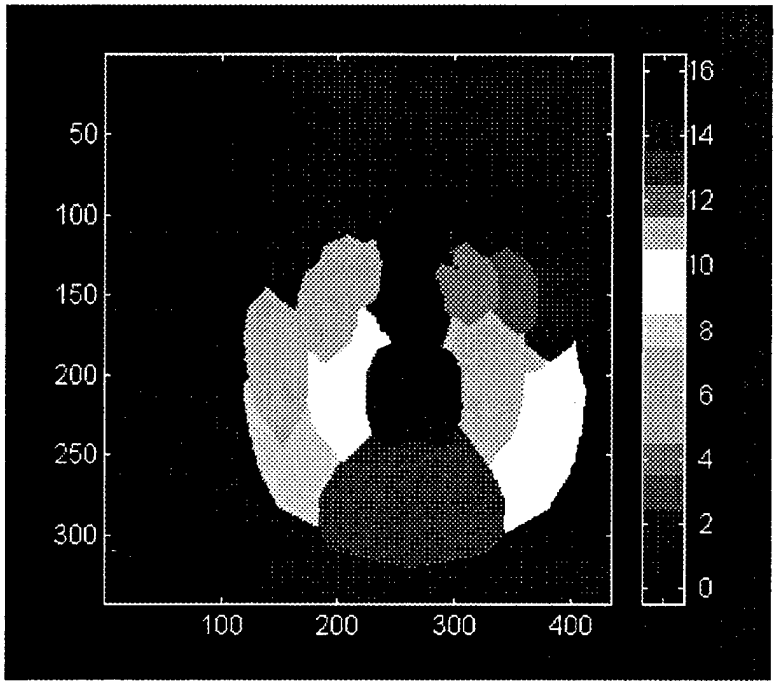


Figure 3.34 - E5424S2I15 - Location of UV pairs for Patch No. 7

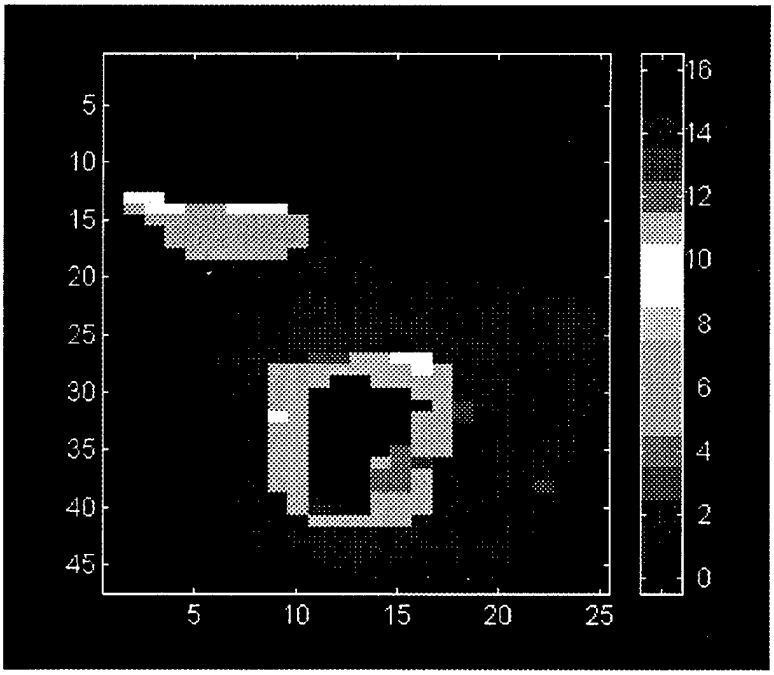


Figure 3.35 - E5424S2I15 - Patch No. 13 with Approximated PDFs

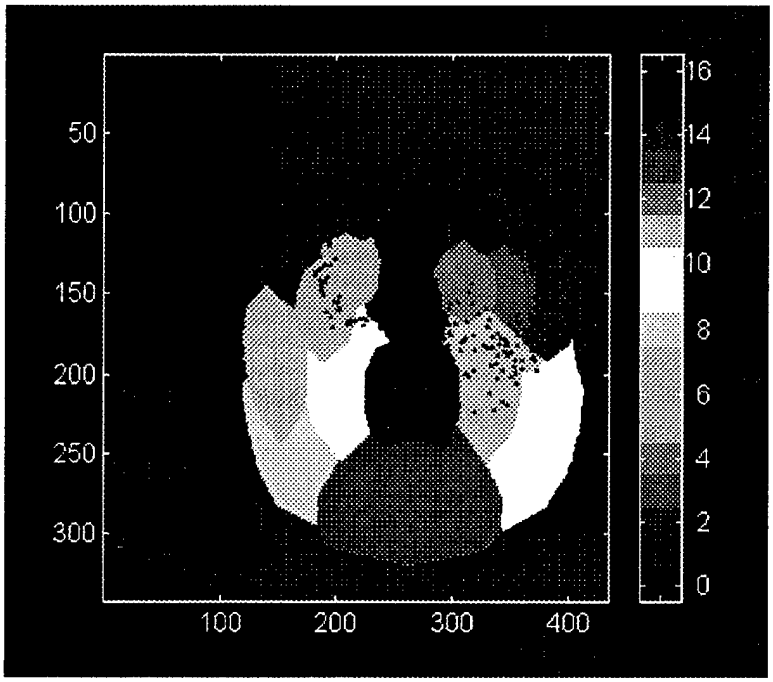


Figure 3.36 - E5424S2I15 - Location of UV pairs for Patch No. 13

Patch No.	1	2	3	4	5	6	7	8	9	10	11	12	13
No. of Pixels	23740	64	30	6	183	1	111	23	1	23	14	49	649
Location wrt Lung	Out	In	In	In	In	In	In	In	In	In	In	In	In

Table 3.7 - E7066S2I17 - Results of labeling

Patch No.	5											
PDF No.	1	2	3	11								
No. of Pixels	19	78	71	15								
Patch No.	7											
PDF No.	4	16										
No. of Pixels	95	16										
Patch No.	13											
PDF No.	1	2	3	4	5	6	9	10	11	13	14	15
No. of Pixels	302	134	6	6	25	5	7	4	61	1	4	94

Table 3.8 - E7066S2I17 - PDFs of the pixels in patches 5, 7, and 13

3.7 – Example 5: Unknown Case Number E18642s2i19

Consider the last case of a lung with a tumor whose type is unknown. The tumor is located in the lower lung of Figure 3.37. Note that the original scene is a 512 by 512 image. As was done in the previous examples, the lung with the tumor is first selected as shown in Figure 3.38. Note that the scene has been reduced to 147 by 174 pixels. The tumor is shown by the patch inside the lung located in the lower middle of the image. Application of the Mapping procedure results in the composite image of Figure 3.39. The component image that contains the tumor is selected as shown in Figure 3.40. The patches in the component image are then labeled to result in the scene of Figure 3.41. Note that, as shown in Table 3.9, patch numbered 4 is the only patch inside the lung that has more than 100 pixels each. Patch 4 is the tumor to be investigated. Table 3.10 summaries the type of approximating PDFs for each pixel in patch 4 as well as the number of pixels with the same approximating PDF.

Processing patch numbered 4 through the Statistical Procedure results in the colored scene of Figure 3.42. Also, Figure 3.43 shows in black the spread of the UV pairs corresponding to the pixels of patch numbered 5. The following are to be noted:

- ❒ In Figure 3.42,
  - ✓ most of the pixels have the same PDF, no contiguous or concentric regions are observed
- ❒ In Figure 3.43,
  - ✓ UV pairs are located on the middle part of the chart
  - ✓ UV pairs are wide spread
  - ✓ UV pairs are spread vertically

With respect to rules 1 to 3 of Section 3.3, it is noted that

(1) the UV pairs are located on the middle part of the chart and not the right part. On the other hand, the UV pairs are wide spread. Note that here, only part of rule 1 is verified. Thus, **no conclusion can be made about whether the patch is a tumor**. Also,

(2) the colored patch does not display any contiguous or concentric behavior of colors. On the other hand, the UV pairs are vertically spread. Thus, only part of rule 2 for a benign tumor is verified. As a result, **no conclusion can be made about whether the patch is a benign tumor**.

Note that in this example that because only parts of rules 1 and 2 apply, it was not possible to get a definite conclusion about whether the patch is a tumor and whether it is Benign.

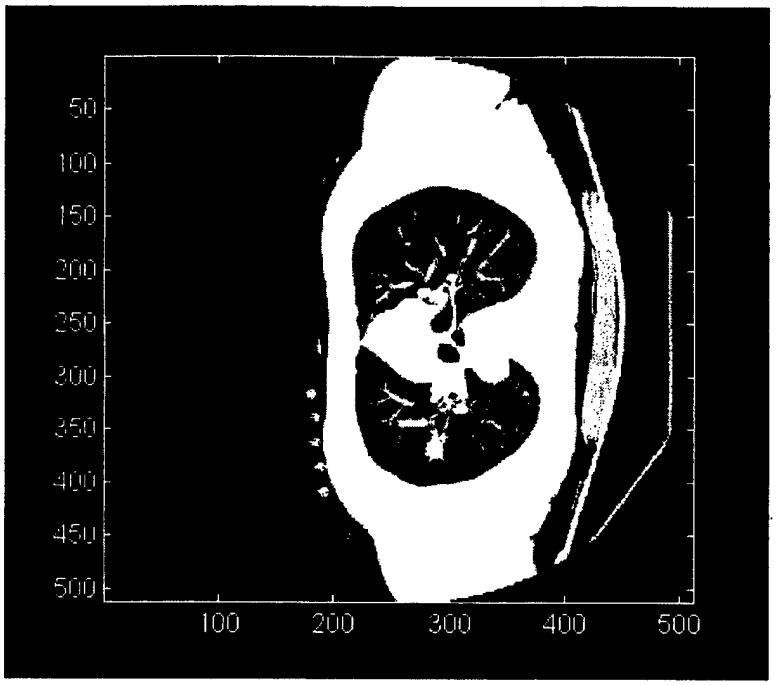


Figure 3.37 - E18642s2i19 - Original Scene

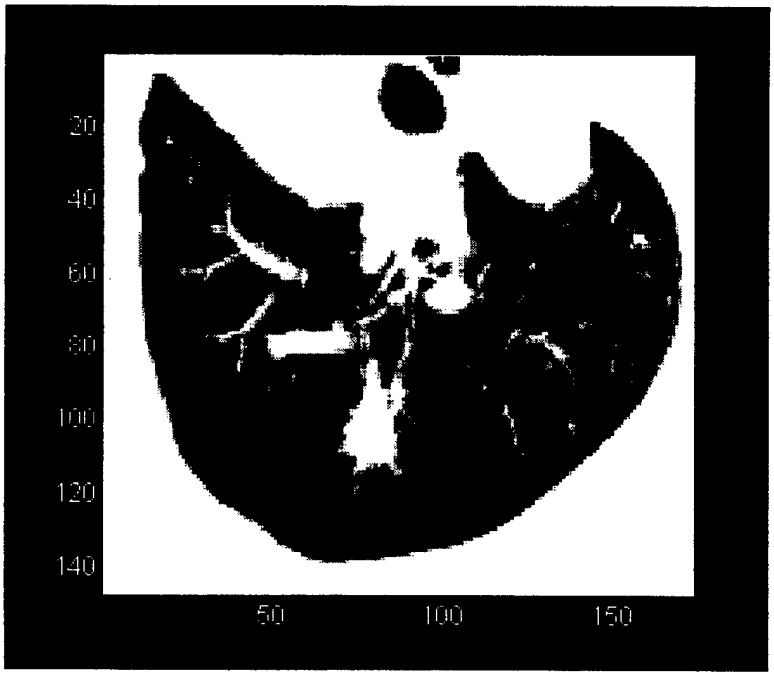


Figure 3.38 - E18642s2i19 - Lung with Tumor

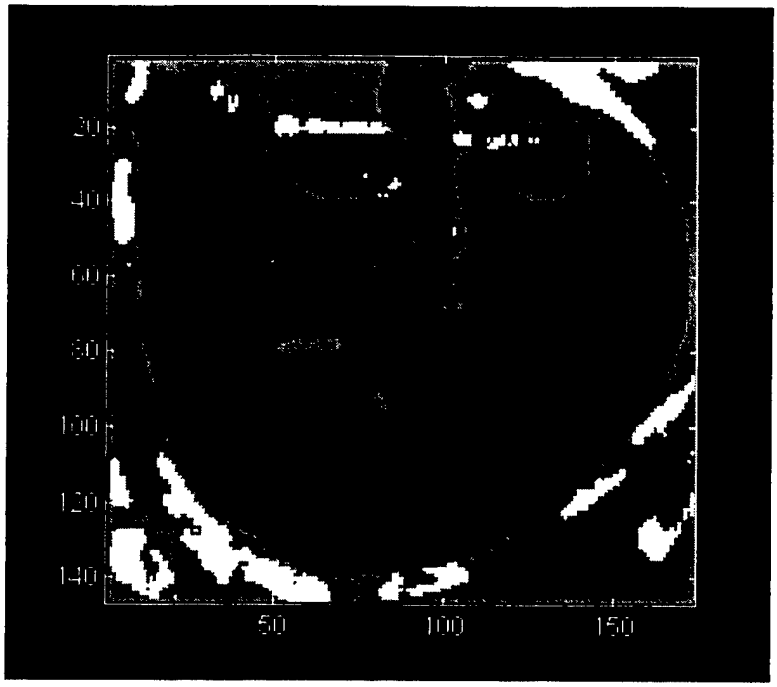


Figure 3.39 - E18642s2i19 - Composite scene

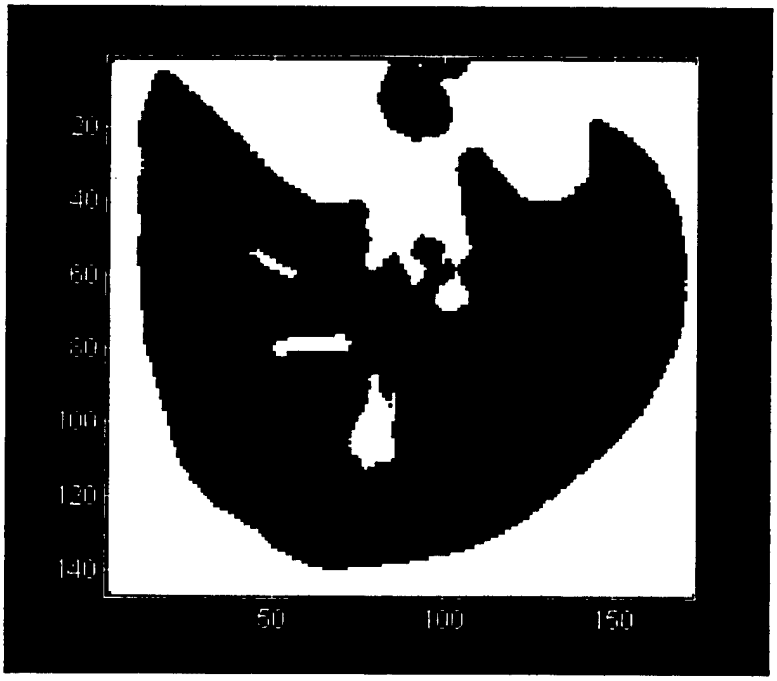


Figure 3.40 - E18642s2i19 - Component Scene Containing Tumor

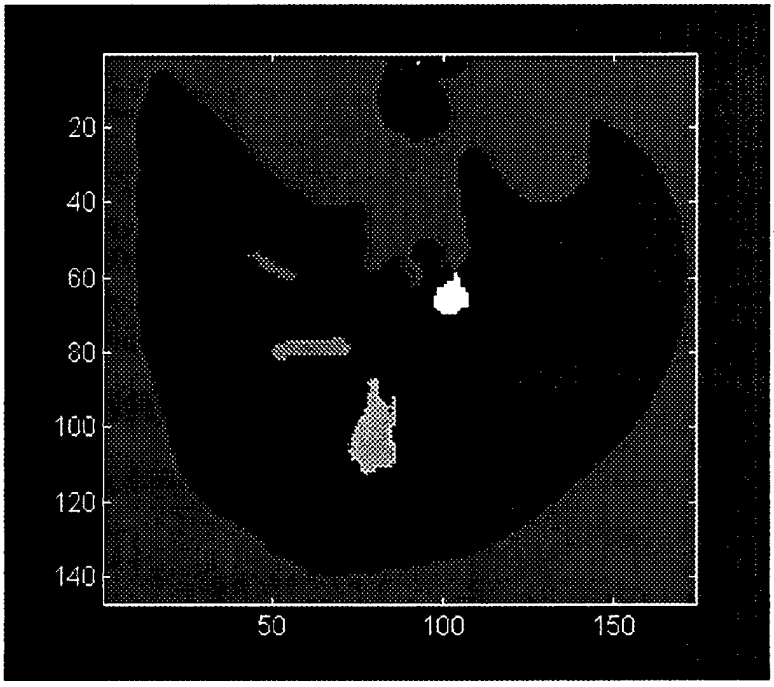


Figure 3.41 - E18642s2i19 - Labeled Scene

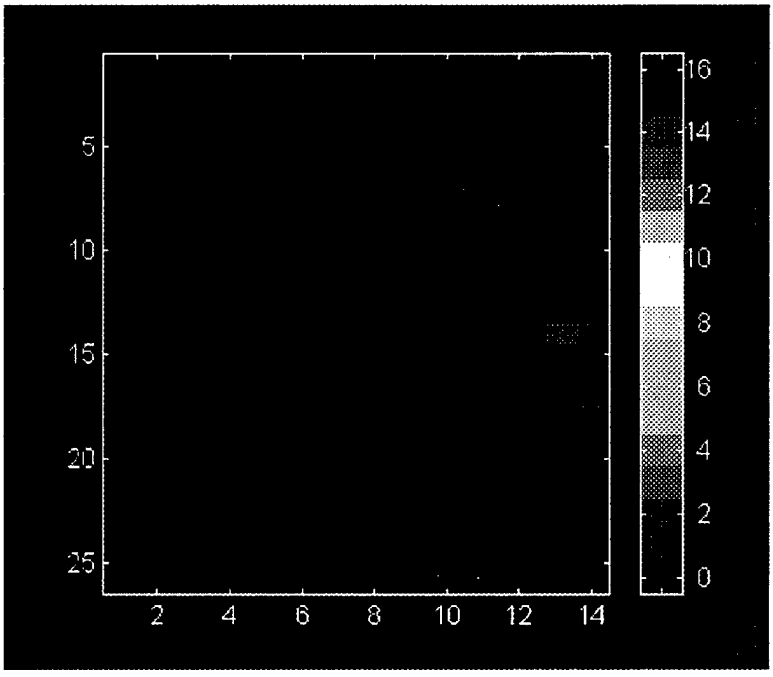


Figure 3.42 - E18642s2i19 -Patch No. 4 with Approximated PDFs

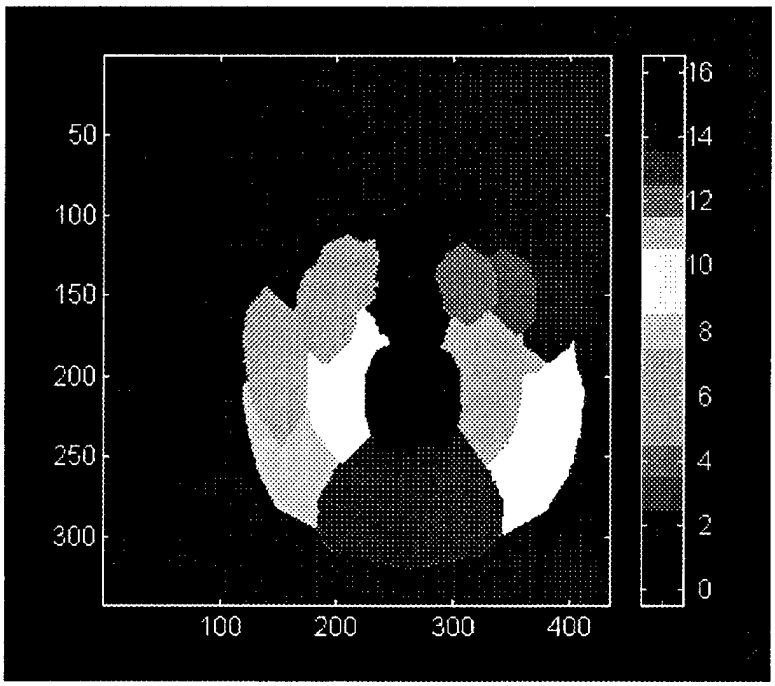


Figure 3.43 - E18642s2i19 - Location of UV pairs of the Patch No. 4

Patch No.	1	2	3	4	5	6
No. of Pixels	10421	33	89	218	1	71
Location wrt Lung	Out	In	In	In	In	In

Table 3.9 - E18642s2i19 - Results of labeling

Patch No.	4		
PDF No.	14	15	16
No. of Pixels	1	204	13

Table 3.10 - E18642s2i19 - PDFs of the pixels in patch 4

Chapter IV

Conclusion

4.1 - Introduction

Recall that the Goals of the research are to,

- detect Tumors located in lungs,
- classify Tumors as Malignant or Benign,
- detect Tumors at very early stages .

In this part of the research the two first goals were particularly targeted. The approach consisted of using (1) using the Mapping Procedure of A'SCAPE to partition and clean the scene, and (2) the Statistical Procedure of A'SCAPE to classify a patch as a tumor or not (detection), and, if it is a tumor, to determine if it is Malignant or Benign (classification).

4.2 – Summary of the results

Using only one sample of a known Benign case and another sample of a known Malignant case, rules were built to detect and classify tumors. The rules were applied of a limited test case library of 3 samples. The results were presented at the Cornell Medical Center and the following answers were provided to the questions raised by this research:

(1) The Mapping procedure is doing a good job by cleaning and "discarding" unneeded information such as branches.

(2) The Mapping procedure needs to keep detailed information about the edges of the patches left in the scene. Note that one of the discriminates that

radiologists use to classify tumors as Benign or Malignant is how smooth or sharp are the edges.

(3) The composite image of a given scene provides more information about the scene then the raw image.

(4) Lung details can be obtained with higher resolution by zooming on regions of interest during CT scanning. This, particularly, solves the problem of the limitation in size of the patches where a minimum of 100 samples are needed by this study to enable the functioning of the Statistical procedure.

(5) A larger library will be provided for rules-building and rules-testing. Namely, a minimum of 20 cases will be provided for rules-building and another minimum of 20 cases for rules-testing.

(6) Instead of trying to classify tumors as Benign or Malignant, it will be more profitable for the radiologists to provide them with "descriptors" that would strengthen their decisions. Such descriptors can be the rules built in Section 3.4. This is because classifying a tumor is a difficult and sometimes an unsolvable problem. As an example, radiologists can look at the tumor to the cell level without being able to decide whether it is Benign or Malignant.

Study of this problem will continue once additional data is received. In addition, the goals will be tailored according to the notes above, and cases of early stages of tumors will be studied.

References

[1] M. A. Slamani, V. Vannicola, D. D. Weiner, "An Automated Approach to the Partitioning and Statistical Characterization of a Surveillance Volume," Proceedings of the 1995 International Conference on Signal Processing Applications & Technology, Boston, MA, 24-26 October, 1995.

[2] M. A. Slamani, V. Vannicola, D. Weiner, "Application of A'SCAPE to the Millimeter Wave Data for the Detection of Concealed Weapons," Proceedings of the International Conference on Millimeter and Submillimeter Waves and Applications III, International Symposium on Optical Science , SPIE'96, 2842-56, Denver CO, 4-9 August 1996.

[3] M. A. Slamani, V. Vannicola, D. Weiner, D. Ferris, "New Statistical Procedure for the Segmentation of Contiguous Nonhomogeneous Regions Based on the Ozturk Algorithm," Proceedings of the Statistical and Stochastic Methods for Image Processing, International Symposium on Optical Science , SPIE'96, 2823-23, Denver CO, 4-9 August 1996.

[4] A. Ozturk and E. J. Dudewicz, "A New Statistical Goodness of Fit Test Based on Graphical Representation," The Biomedical Journal, 1991

MISSION
OF
AFRL/INFORMATION DIRECTORATE (IF)

The advancement and application of information systems science and technology for aerospace command and control and its transition to air, space, and ground systems to meet customer needs in the areas of Global Awareness, Dynamic Planning and Execution, and Global Information Exchange is the focus of this AFRL organization. The directorate's areas of investigation include a broad spectrum of information and fusion, communication, collaborative environment and modeling and simulation, defensive information warfare, and intelligent information systems technologies.